Genetic variation and complex rearrangements in Autism Spectrum Disorders: implications for genetic counseling

Marta Codina i Solà

TESI DOCTORAL UPF / 2016

THESIS SUPERVISORS

Prof. Luis A. Pérez Jurado
Dr. Ivon Cuscó Martí

Departament de Ciències Experimentals i de la Salut

Universitat Pompeu Fabra Barcelona

La sirena

Tenia el cabàs penjat darrere la porta
i a les mans el solc que hi va deixar l'eina.
Però els matins s'allarguen
i del llit estant, amb els ulls oberts,
sentim com toca la sirena.
Si comptem amb els dits, ens sobra temps;
Si comptem amb diners ens manca feina.
Però encara hem de viure
i del llit estant, amb els ulls oberts,
sentim com toca la sirena.
No és pas culpa vostra, ens diu sovint la gent,
aquests temps són durs i tot escasseja.
Però els anys no s'aturen
i del llit estant, amb els ulls oberts,
sentim com toca la sirena.
Vindran altres temps si ho volem tots plegats;
mai ens han fet por l'esforç ni la feina,
i tant se val si ja hem mort quan ningú
del llit estant, amb els ulls oberts,
senti com toca la sirena.

Miquel Martí i Pol

Als meus pares,

a la Núria,

a en Gerard

Agraïments

Durant aquests anys de doctorat, he tingut la gran sort de conèixer i d'aprendre de moltíssima gent diferent. Sense ells, sense vosaltres, res d'això no hauria estat possible. La llista és molt llarga però, com tots sabeu, la meva memòria no gaire, així que espero no deixar-me ningú.

En primer lloc, m'agradaria agrair al Luis i a la Ivon la seva dedicació i orientació durant tots aquests anys. Al Luis, gràcies per la doble oportunitat de fer el doctorat i el màster d'assessorament genètic i tot el que això m'ha permès aprendre sobre la genètica al llarg d'aquests anys. Ivon, et voldria agrair no només el que m'has ensenyat sobre autisme, genètica o tècniques moleculars, sinó també tot el m'has demostrat dia a dia, la teva capacitat de treball en equip, el teu recolzament constant i la teva generositat.

Moltes gràcies, també, a la gent que formeu o que heu format part del laboratori 422 per haver compartit tantes estones al llarg d'aquests anys: Clara, Raquel, Roser, Tina, Aïda, Gaby, Judith (aquí i al màster!), Mariví, Maria, Jairo, Anna, Armand, Débora, Marcos, Francesc, Guillermo, Eulàlia, Arturo, Paula, Gaudi, Marina, Mireia, Ricard i tanta i tanta altra gent. Aquests anys no haguessin estat el mateix sense el vòlei (i les festes), les copes (encara que siguin robades), els dinars de sushi, les braves o les nits de jocs!

Tina, t'haig de donar les gràcies per tantes i tantes coses, que ara sóc incapaç de resumir-les. Primer de tot, per ensenyar-me a relativitzar-ho tot, incloent MLPAs. Per escoltar-me sempre, per fer-me costat i, també, per saber-me parar els peus quan tocava. Quina sort he tingut de començar i acabar aquest camí al teu costat! Entre totes dues hem aconseguit arribar a la recta final sense acabar gaire més boges del que estàvem en començar. Els atacs de riure que ens agafen de tan en tan em sembla que ja no els podrem arreglar, però alguna seqüela ens havia de quedar!

Roser, gràcies per estar sempre disposada a ajudar i per tantes estones compartides al gimnàs, corrent, fent bici, però sobretot xerrant! També pel teu suport incondicional, pel teu bon caràcter (sí, malgrat tot el que crides!) i per haver-nos mimat tant a mi i a la Tina! *Vamos, equipo*!

Aïda, gràcies per ser la companya de taula més dolça, detallista, innocent i pacient que em podria haver tocat mai! Gràcies per ensenyar-me a fer gràfics amb l'R, i per tants moments al teu costat i al d'en Martí i en Marcel!

Gaby, gràcies per compartir els trajectes en bus i tantes i tantes altres coses! Pel teu suport i els teus consells, per motivar-me perquè jugués a vòlei i per la teva paciència per intentar que la pilota passés a l'altra banda de la xarxa!

Clara, gràcies per les xerrades, les xafarderies, el teu riure que traspassa parets, les bromes que fas i les que et fan! També pels teus consells, per la supervisió treball de màster i les consultes i tot el que he après gràcies a tu!

Raquel, gràcies, per la teva ajuda aquests últims mesos i per intentar que em deixessin imprimir en el meu moment de crisi! També per fer-nos riure en tants moments i per haver compartit tantes coses més enllà de la ciència!

També vull donar les gràcies a tot l'equip de qGenomics per compartir laboratori, sopars de Nadal i amics invisibles: Manel, Cristina, Sònia, Olaya, Lluís, Xavi, Ma José, Sílvia i tota la resta. I, sobretot, Benja, gràcies per la teva paciència i ajuda en els meus inicis amb la pantalla negra i també per tots els favors que t'he demanat!

A més de la gent del laboratori, també he tingut la sort de conèixer una colla d'assessors genètics fantàstics que han fet que fer el màster d'assessorament fos una experiència única: Adrià, Albert, Anna, Diana, Èlia, Gisela, Gina, Josep, Judith, Luna i Mònica. Moltíssimes gràcies a tots! Diana i Gisela, gràcies per tot el que vam riure i compartir durant les pràctiques a Vall Hebron! Adrià, gràcies per ser el dimoniet que xiuxiueja darrera l'orella esquerra i Diana, a tu per ser l'àngel de l'orella dreta! Sou genials!

Gràcies Èrica, Martí, Helena, Ari, Berta, Helena (Palumbo) per haver-me fet costat tot aquest temps, per ajudar-me a desconnectar tots els caps de setmana i per ser-hi sempre! Helena, Mia i Alba, gràcies per tot el temps compartit a la uni, pràctiques, dinars amb carmanyola als barracons i, sobretot, pel tots els moments que vam passar fora!

Gerard, gràcies pel teu suport incondicional, per la portada que vindrà i per tants matins culturals! Aquests últims dies has tingut una paciència immensa, però, com sempre i encara no sé com, has fet que tot sigui fàcil i que sembli encara més senzill.

Gràcies també a la meva família, tan petita i tan valuosa, que cap sencera dins d'uns agraïments. Als pares, gràcies per haver-nos inculcat, a mi i a la Núria, la importància d'estudiar allò que ens agradava, que no deixa de ser res més que aprendre. Gràcies, també, per transmetre'ns que no és només important allò que dius, sinó, també, com ho dius. Núria, gràcies per ser-hi sempre i per saber-me acompanyar des d'aquí i des de la distància. Avi Ramon, gràcies per la teva fermesa i per preocupar-te sempre per nosaltres. Iaia Margarita, gràcies per l'alegria i la vitalitat que vas saber-nos transmetre fins l'últim moment i que mai oblidaré. Tiets, gràcies pels ànims durant aquests últim temps i per no empipar-me tan com normalment! Iaia Carme, gràcies per rebre'm amb tanta alegria els divendres, per ser una *rara avis* i donar-me un motiu per a presumir de la meva pròpia genètica!

Moltes gràcies a tots!

Mi diagnóstico es sencillo, sé que no tengo remedio

Julio Cortázar

Abstract

The etiology of Autism Spectrum Disorders (ASD) remains unknown for most of the cases, in spite of its strong genetic component. A greater knowledge of its genetic basis would result in many benefits, including specific genetic counseling for families and, eventually, the development of personalized therapeutic strategies. In this thesis, we have applied several recent sequencing technologies and adapted pipelines to its study. We have investigated the role of rare variants and its transcriptional consequences and explored the contribution of complex rearrangements to its missing heritability. In addition, we have studied second-hit susceptibility genetic factors in a group of individuals with Williams-Beuren syndrome, a genomic disorder associated with a mirror phenotype. Finally, we have explored parental knowledge and the effect of genetic counseling in affected families. Our results reveal that both highly penetrant mutations and inherited variants of milder effect contribute to its susceptibility, following monogenic and oligogenic or multifactorial models, respectively. Each of them may contribute to part of the variability and may explain a subset of cases.

Resum

Malgrat el fort component genètic dels Trastorns de l'Espectre Autista (TEA), l'etiologia de la majoria de casos es desconeix. Un major coneixement de les seves bases genètiques seria molt beneficiós, ja que permetria un assessorament genètic específic a les famílies i, a la llarga, el desenvolupament d'estratègies terapèutiques personalitzades. En aquesta tesis, s'han aplicat diverses tècniques recents de seqüenciació i estratègies d'anàlisi adaptades. S'ha investigat el paper de variants rares i les seves conseqüències transcripcionals, així com de reordenaments complexos. A més, hem estudiat la presència de variants de susceptibilitat en un grup de persones amb síndrome de Williams, un trastorn genòmic associat a un fenotip oposat. Finalment, hem explorat el coneixement i les opinions en un grup de famílies afectades i, també, l'efecte de l'assessorament genètic. Els resultats obtinguts indiquen que, en el TEA, hi contribueixen tant mutacions altament penetrants, com variants heretades que augmenten lleugerament el risc i poden seguir tant models monogènics com oligogènics. Cada un d'aquests models contribuiria a explicar part de la variabilitat i dels casos.

Prologue

The development of new sequencing technologies in the recent years has revolutionized our understanding of genetics and its relation to disease. Autism Spectrum Disorders (ASD) have not been led behind in this quest. The application of these new techniques has revealed the implication of hundreds of new loci and the great genetic heterogeneity behind its huge clinical diversity.

This thesis expands the knowledge about the genetic basis of ASD, relying on the application of various next-generation sequencing technologies, including exome and transcriptome sequencing. In addition, it explores knowledge and perspectives in affected families and evaluates the utility of genetic counseling. The thesis is structured as follows.

The introduction gives a general overview of the clinical characteristics, molecular basis and medical management of ASD, and also addresses the role of segmental duplications in disease and evolution, focusing on the 7q11.23 region and its relation to neurodevelopmental disorders. The last part provides a general picture about genetic counseling and specifically focuses on genetic counseling in idiopathic ASD.

The main body of the thesis is divided in five chapters corresponding to different studies that describe in detail the methods and results obtained.

The discussion aims to integrate and interpret the results obtained in previous chapters, as well as to place them into the framework of previous knowledge of the genetic basis of ASD. Following the discussions, the main findings of the thesis are summarized in the conclusions.

Contents

List of figures

Introduction

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Discussion

List of tables

Introduction

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of abbreviations

AAP	American Academy of Pediatrics
ABA	Applied Behavior Analysis
aCGH	Array Comparative Genomics
ACMG	American College of Medical Genetics and Genomics
ADDM	Autism and Developmental Disabilities Monitoring
ADHD	Attention Deficit Hyperactivity Disorder
ADI-R	Autism Diagnostic Interview- Revised
ADOS	Autism Diagnostic Observation Schedule
ASD	Autism Spectrum Disorders
ASE	Allele-Specific Expression
BAP	Broader Autistic Phenotype
CDC	Center for Disease Control and Prevention
CHARGE	Childhood **Autism** Risks from Genetics and the Environment
CHAT	Checklist for Autism in Toddlers
CMA	Chromosomal Microarray
CNVs	Copy Number Variants
CPDB	ConsensusPathDB
DAS	Differential Ability Scale
DGV	Database of Genomic Variants
DP	Depth of Coverage
DSM	Diagnostic and Statistical Manual of Mental Disorders
EEG	Electroencephalogram
EVS	Exome Variant Server
FDR	False Discovery Rate
FISH	Fluorescence in Situ Hybridization
FoSTeS	Fork Stalling and Template Swithcing
FPKM	Fragments per Kilobase of Transcript per Million mapped reads
GATK	Genome Analysis Toolkit
GPRD	General Practice Research Database
GTEx	Genotype-Tissue Expression
GWAs	Genome Wide Association Studies
ICD	International Classification of Diseases
ID	Intellectual Disability

IGV	Integrative Genome Viewer
IQ	Intelligence Quotient
ISCA	International Standard Cytogenomic Array
ISSX	X-linked infantile spasm syndrome
LCR	Low Copy Repeats
LGD	Likely Gene Disruption
LoF	Loss of Function Variants
M-CHAT	Modified Checklist for Autism in Toddlers
MAF	Minimum Allele Frequency
MLPA	Multiple Ligation Probe Amplification
MRI	Magnetic Resonance Imaging
NAHR	Non-allelic Homologous Recombination
NETBAG	NETwork-Based Analysis of Genomic variations
NGS	Next Generation Sequencing
NHEJ	Non Homologous End Joining
NSGC	National Society of Genetic Counselors
OCD	Obsessive-compulsive Disorder
PAR1	Pseudoautosomal Region 1
PBMCs	Peripheral Blood Mononuclear Cell
PDD-NOS	Pervasive Developmental Disorder-Not Otherwise Specified
PDDST-II	Pervasive Developmental Disorders Screening Test-II
PECS	Picture Exchange Communication System
PHTS	PTEN hamartoma tumor syndrome
PPI	Protein Protein Interaction
PSV	Paralogous Sequence Variant
RR	Recurrence Risk
RPKM	Reads per Kilobase per Million mapped reads
SCQ	Social Communication Questionnaire
SD	Segmental Duplications
SLI	Specific Language Impairment
SNP	Single Nucleotide Polymorphisms
SNV	Single Nucleotide Variant
SPARK	Simons Foundation Powering Autism Research for Knowledge
SSC	Simons Simplex Collection
SSRIs	Serotonin-specific reuptake inhibitors

SVAS	Supravalvular Aortic Stenosis
TEACCH	Treatment and Education of Autistic and related Communication Handicapped Children
VOUS	Variant of Unknown Significance
WBS	Williams-Beuren syndrome
WES	Whole-Exome Sequencing
WGCNA	Weighted Correlation Network Analysis
WGS	Whole Genome Sequencing
WHO	World Health Organization
XCI	X chromosome inactivation
ZRPKM	Standarized Z Reads per Kilobase per Million mapped reads -scores of Reads per Kilobase per Million mapped reads

INTRODUCTION

1. Autism Spectrum Disorders

Autism Spectrum Disorders (ASD) [Online Mendelian Inheritance in Man (OMIM) 209850] are a group of neurodevelopmental disorders characterized by alterations in three main domains: communication deficits, impaired social interaction and a restrictive and repetitive pattern of behavior. Dr. Leo Kanner was the first to describe eleven children with “infantile autism” in a paper entitled “Autistic Disturbances of Affective Contact” in 1943 (1). One year later, Dr. Hans Asperger described four males with normal intelligence but "a lack of empathy, little ability to form friendships, one-sided conversation, intense absorption in a special interest, and clumsy movements" (2).

1.1 Clinical features

ASD comprise a heterogeneous group of complex disorders, which manifestations can vary greatly among individuals. The first signs and symptoms, such as a lack of eye contact, not sharing things with others or an strict adherence to routines, usually appear from early childhood, typically before three years old (3). About 30% of children with ASD may suffer a regression of the previously acquired skills, usually after the first or second year of age. Although language is probably the most frequently lost skill, a substantial proportion of children loose social skills or both (4).

Regarding communication deficits, language impairments can range from a complete lack of speech in some individuals, delays in language acquisition in others or deficits in comprehensive speech or echolalia. Although some individuals may present intact formal language, social communication is still impaired. For example, language may be used in an overly literal manner and not be adequate to the social situation. Nonverbal communication is also affected in individuals with ASD. For instance, verbal and nonverbal communication may be poorly integrated or individuals may display alterations in eye contact, difficulties to understand gestures or deficits in the use of facial expressions (5,6).

In the social interaction area, many affected individuals show a reduced or a lack of social interest. Young children may show deficits in social and imaginative play and adults may display an inappropriate behavior or a preference for solitary activities. As an example, young children may not share objects such as toys with their parents and, although they may initiate

physical contact, they may use their hands or arms as tools rather than seeking their company (5,6).

Restricted and repetitive patterns of behavior, interests or activities include motor stereotypies or repetitive patterns of movement, echolalia, resistance to change or strict adherence to routines. In this regard, children with ASD may experience stress at small changes such as varying the road taken to school or manifest an obsession for particular colors or shapes (5,6).

ASD are frequently associated with concurrent medical, developmental or psychiatric conditions. Regarding comorbid neurodevelopmental disorders, Intellectual disability (ID) is present in approximately 45% of individuals with ASD, although this estimate may vary if verbal skills are considered. ID is a strong predictor of poor prognosis and may be the most common co-occurring disorder with ASD (7,8). Attention-deficit hyperactivity disorder (ADHD) is also frequent among children with ASD (28-44%), as well as tic disorders (14–38%) and motor abnormalities (<80%) (7).

Medical conditions such as epilepsy, gastrointestinal problems and sleep abnormalities are also more prevalent among children with ASD. Individuals with ID or genetic syndromes present higher rates of epilepsy and seizures are a prognostic factor of poor outcome. Its incidence peaks in early childhood and in adolescence (7,9). Gastrointestinal problems in children with ASD include constipation, gastro-esophageal reflux, abdominal pain and diarrhea. They are estimated to affect between up to 70% of children (rang extension 9-70%) and are a frequent concern among parents (7,10,11). Sleep disorders are also frequent and are estimated to affect around 80% of individuals, mainly insomnia (7).

In addition, psychiatric disorders such as anxiety, depression, obsessive-compulsive disorder (OCD) and psychotic disorders are also common among individuals with ASD (7,12,13). An important proportion of children with ASD also present macro or microcephaly. Macrocephaly, defined as head circumference >97th centile is present in about 16% of children with ASD, whereas microcephaly (head circumference <3rd centile) is found in a similar proportion of patients, approximately 15% (14,15). Other minor and major physical anomalies are also more frequent in ASD patients than in general population (16).

1.2 Prevalence

Estimates of prevalence of ASD vary greatly among studies and in time. In the United States it was estimated to affect 1 in 68 children in 2010 according to the Center for Disease Control and Prevention (CDC) Autism and Developmental Disabilities Monitoring (ADDM) Network (17). As it can be seen in Table 1, the CDC reported approximately a twofold increase in prevalence from 2000 to 2008. It has been hypothesized that this increase may be due to higher awareness, a shift in diagnostic criteria or environmental causes. Some authors have proposed that it may be partly due to a shift in the type of diagnosis, particularly to a decrease in diagnosis of ID (18,19).

Prompted by the press impact caused by the release of the CDC's data, Taylor et al (20), conducted a similar and comparable study based on the UK General Practice Research Database (GPRD). The study focused on 8-year-old children to allow for comparison and annual prevalence estimates were calculated for boys and girls separately. In 2010, the prevalence rate for boys aged 8 in the UK was approximately 1 in 256, a considerable lower figure compared to the estimates of 1 in 68 children or 1 in 42 boys in the CDCs' data. Moreover, although both in the UK and USA the prevalence increased over the 90s (21), prevalence and incidence have remained stable in the UK since 2004 (Table 1). Therefore, the rise in prevalence of ASD over the last years remains a controversial issue and it is still unclear whether it may be due to underlying environmental factors or methodological concerns.

Surveillance Year	Birth Year	CDC prevalence rate per 1000	GPRD prevalence rate per 100
2000	1992	6.7 (4.5—9.9)	NA
2002	1994	6.6 (3.3—10.6)	NA
2004	1996	8.0 (4.6—9.8)	3.58 (3.28–3.8)
2006	1998	9.0 (4.2—12.1)	3.86 (3.56–4.19)
2008	2000	11.3 (4.8—21.2)	3.91 (3.59–4.25)
2010	2002	14.7 (5.7—21.9)	3.9 (3.57–4.24)

Table 1. Prevalence estimates of ASD in the United States and UK from 2000 to 2010 according to (17,20,22–25). CDC data represents prevalence in all children, whereas GPRD prevalence rates include only boys. GPRD data from 2000 to 2002 is not available.

However, both studies agree on the fact that ASD are five times more prevalent among men than women, although this ratio lowers to one male for every female among the group of patients with more severe manifestations (17).

1.3 Diagnostic criteria

Various classification systems for mental health disorders have been developed over the years. Currently, the most used by mental health professionals are the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) by the American Psychiatric Association (2013) (5) and the 10th edition of the International Classification of Diseases (ICD) (26) by the World Health Organization (WHO).

ASD definition and diagnostic criteria have varied greatly in the editions of the DSM. The third edition of the DSM (DSM-III) was the first to establish ASD as distinct diagnostic category. It was then termed “Infantile Autism” and was based in 6 characteristics, with an onset before 30 months of age which included lack of responsiveness to other people, deficits in language development and resistance to change or peculiar interests (27).

The fourth edition of the DSM in 1994 and its revised edition in 2000 (DSM-IV and DSM-IV-TR) (28,29) established several subtypes of autistic disorders: Autistic Disorder, Asperger disorder, Childhood disintegrative disorder and Pervasive developmental disorder not otherwise specified (PDD-NOS). In addition, the triad of features in which the current definition of ASD is based was established: impairments in social interaction, impairments in communication and a restricted, repetitive and stereotyped pattern of behavior, interests and activities. A total of 16 symptoms were listed as present in ASD, although only six were needed for a diagnosis of autism (Table 2). The latest and current edition of DSM (DSM–5), published in 2013, unified the previous categories into the term ASD and reorganized the classical triad of features into two: difficulties in social communication and social interaction and restricted and repetitive behavior, interests, or activities (5).

		DSM-IV	DSM-5	ICD-10
CATEGORIES		- Autistic Disorder - Asperger disorder - Childhood disintegrative disorder - PPD-NOS		- Childhood autism - Atypical autism - Rett syndrome - Asperger syndrome - Unspecified pervasive developmental disorder - Overactive disorder associated with mental retardation and stereotyped movements - Other childhood disintegrative disorder
SOCIAL		(2): - Multiple nonverbal inadequate behaviors - Failure to develop peer relationships - Lack of sharing with other people lack of social reciprocity	All: - Impairment social and emotional reciprocity - Inadequate nonverbal communicative behaviors - difficulties in developing and maintaining relationships	(1): - Failure adequately to use non- verbal body language - Failure to develop peer relationships - Lack of socio-emotional reciprocity and modulation
COMUNICATION		(1): - Delay or lack of spoken language - if speech is present, impairment to initiate and sustain a conversation - Stereotyped and repetitive use of language - Limited imitative play		(2): - Delay or lack of spoken language - Failure to initiate maintain a conversational interchange - Stereotyped and repetitive use of language - Abnormalities in speech (pitch, stress, rate, rhythm and intonation)
BEHAVIOUR		(1): - Stereotyped and restricted patterns of interest - Inflexible with non-functional routines or rituals - Stereotyped and repetitive motor mannerisms - Preoccupation with parts of objects	(2): - Stereotyped or repetitive speech, motor movements, or use of objects - Adherence to routines, ritualized patterns of verbal or nonverbal behavior, or excessive resistance to change - Restricted, fixated interests - Hyper/hypo-reactivity to sensory input or unusual interest in sensory aspects	(2): - Stereotyped and restricted patterns of interest - Compulsive adherence to specific, non-functional, routines or rituals - Stereotyped and repetitive motor mannerisms - Preoccupations with part-objects or non-functional elements - Distress over changes in small, non-functional, details of the environment

Table 2. Differences in diagnostic criteria and classification of ASD between DSM-IV, DSM-5 and ICD-10.

In this respect, the classification and definition of ASD in the 10th Revision of the International Statistical Classification of Diseases and Related Health Problems (ICD-10), is similar to that of the DSM-IV. The ICD-10 includes ASD under the term PDD, which are defined as a group of disorders characterized by the classic triad of features mentioned above. Moreover, the ICD-10, provides eight different categories to further classify the disorder: Childhood autism, Atypical autism, Rett syndrome, Asperger syndrome, unspecified PDD, overactive disorder associated with mental retardation and stereotyped movements and other childhood disintegrative disorder. For a summary and comparison of criteria used between DSM-IV, DSM-5 and ICD-10, see Table 2.

It is still debated how changes in DSM-5 and DSM-IV-TR will affect individuals diagnosed with ASD. Some studies have pointed out that changes between editions may exclude high functioning atypical cases from diagnosis, difficult early diagnosis and result in lower prevalence estimates. This may affect services provided to families and patients (30–35).

1.4 Genetic bases of ASD

ASD are thought to have a strong genetic component. This is supported by heritability studies comparing affected status in monozygotic and dizygotic twin pairs. Several studies across the years have shown higher concordance between monozygotic twin pairs (36-96%) than between dizygotic pairs (0-31%) (36–43). In addition, recurrence risk for siblings of an affected child is estimated to be between 5-10% for classic autism and 20% for a broader autistic phenotype (BAP), much greater than prevalence estimates in the general population. Recurrence risk in families with two affected children is even higher and is estimated to be around 30% (44). The relative risk of ASD is also increased in extended relatives of affected members, although it decreases as the degree of familial relationship diminishes. For full siblings, the relative risk compared to the general population is 10.3 (95% CI: 9.4–11.3), for maternal and paternal half siblings is 3.3 (95% CI: 2.6–4.2) and 2.9 (95% CI: 2.2–3.7) respectively, and for cousins, 2.0 (95% CI: 1.8–2.2) (45). Therefore, given its high heritability (36-96%), ASD may be the most heritable of developmental disorders. As a result, a large number of studies have tried to identify the underlying genetic causes.

Despite the body of evidence showing that genetics play a major role in ASD, currently a genetic cause can be identified in only 20 to 25% of cases (46). For the rest of cases (75-80%), the etiology remains unknown. The growing knowledge about genetic causes of ASD shows that its clinical heterogeneity is reflected in a high degree of genetic heterogeneity. Genetic factors involved in ASD comprise all type of genetic variation: chromosomal alterations, triplet expansions, *de novo* or rare inherited single nucleotide variants (SNVs) and copy number variants (CNVs) (Figure 1). Therefore, ASD present a high degree of genetic heterogeneity, with various types of genetic variation that can follow several modes of inheritance.

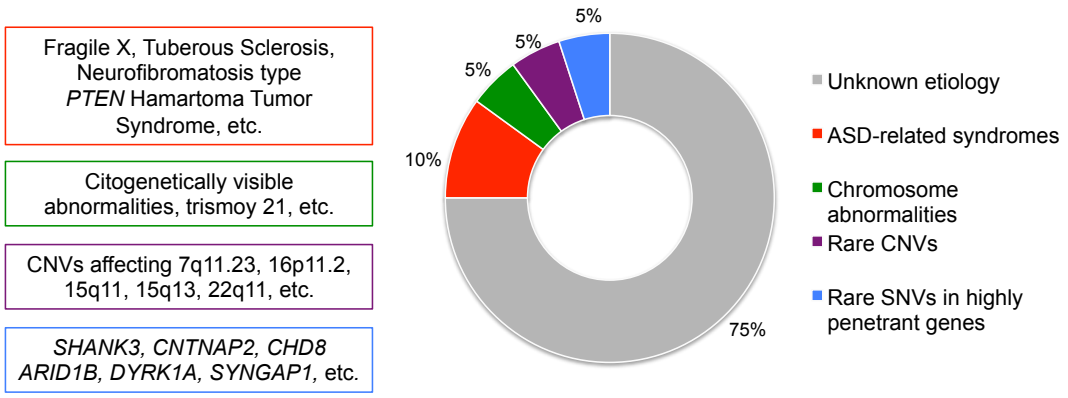


Figure 1. Schematic breakdown of genetic factors contributing to ASD. Adapted from (46).

1.4.1 ASD-related syndromes

Approximately 10% of patients with ASD present a known syndrome, in which autism is one of the comorbid features. Among the disorders with a higher prevalence of ASD symptoms, some of the most known are: X-fragile syndrome (*FMR1*), tuberous sclerosis (*TSC1*, *TSC2*), neurofibromatosis type I (*NF1*), Rett syndrome (*MECP2*) and *PTEN* Hamartoma Tumor Syndrome (PHTS) related disorders (*PTEN*). Other genetic syndromes with prevalent ASD manifestations include Cri du Chat syndrome (5p deletion), the 7q11.23 duplication syndrome or WAGR syndrome (11p13 deletion). More than 100 genes and 44 loci have been proposed as causative genetic factors for ASD-related syndromes, proving the high genetic heterogeneity of the phenotype and its genetic and biological overlap with other disorders such as ID or epilepsy (47).

1.4.2 Chromosome abnormalities

Regarding cytogenetically visible chromosome abnormalities, around 5% of patients with ASD are caused by a chromosomal rearrangement of this type. This type of rearrangements is more common in patients with dysmorphologic features (48). One of the most prevalent abnormalities is a gain of 15q11-q13 either through direct duplication or as a supernumerary marker chromosome, which includes the imprinted ubiquitin-protein ligase (*UBE3A*) gene, responsible for Angelman syndrome. Aneuploidies, which are the presence of an abnormal number of chromosomes, have also been related to ASD. About 7% of patients with Down syndrome (trisomy 21), present ASD and a higher percentage show ritualistic and obsessive behaviors (49). The presence of an extra X chromosome, either in boys (46, XXY Klinefelter syndrome) or girls (XXX, Trisomy X) has also been related to a higher frequency of autistic symptoms (50).

1.4.3 Rare CNVs

Since 2004, the advent of chromosomal microarray (CMA) has enabled the detection of submicroscopic CNVs not detected previously by conventional karyotype (51,52). In the last ten years, its application to the study of disease, and especially neurodevelopmental disorders, has enabled the detection of submicroscopic CNVs in about 10% of ASD patients (53–56). Several studies have shown that the burden of *de novo* CNVs is higher in affected individuals than controls or their healthy siblings (56–58). In a study by Sebat et al 10% of patients with sporadic ASD were found to carry a *de novo* CNV, compared to only 3% of patients with an affected first-degree relative and 1% of controls (56). Similar proportions were described by Sanders et al in a similar study, where 5.6% of probands were found to carry a large *de novo* CNV, compared to 1.7% of siblings (58).

Some of those CNVs are recurrent events, including microduplication or microdeletion syndromes that can be either *de novo* or in some cases inherited from unaffected progenitors, showing incomplete penetrance and variable expressivity. Among these, some of the most frequent are the 7q11.23 duplication, 16p11.2 duplication and deletion, maternal 15q11.2-13.1 duplication, deletion or duplication of 15q13.2-13.3 and deletion or duplication of 22q11.2 (Table 3). Some of these CNVs are implicated in a

wide range of disorders, such as schizophrenia, ID or milder phenotypic characteristics(59).

CNV (deletion and duplication events combined)	Events in cases/controls (n=2120)	Frequency in ASD cases	P-value
16p11.2	18/2120; 3/2159	0,8%	0,001
7q11.23	4/2120; 1/2159	0,2%	0,06
22q11.2	4/2120; 1/2159	0,2%	0,214
1q21.1	4/2120; 3/2159	0,2%	0,723
15q13.3	5/2120; 0/2159	0,2%	0,030
15q11-q13	2/2120; 0/2159	0,1%	0,245

Table 3. Recurrent microdeletion and microduplication CNVs implicated in ASD.

For instance, one of the recurrent deletions affecting the 16p11.2 region (between BP4 and BP5) has been found to be present in approximately 0.6% of patients with ASD but only in 0.04% of controls (60–64). A study characterizing the cognitive and behavioral phenotype of 16p11.2 deletion carriers showed that 90% presented psychiatric or developmental disorders and that those not meeting criteria for ASD presented a higher frequency of autism-related features. Moreover, although individuals showed a high range of intelligence, Intelligence Quotient (IQ) scores among carriers were 26 points lower than those of non-carrier family members (65). This suggests that, for some CNVs and traits, penetrance may be complete, even though expressivity may vary greatly among individuals.

1.4.4 Rare SNVs

Besides comorbid monogenic Mendelian disorders in which autism is one of multiple clinical features, SNVs in highly penetrant genes have also been described as causes of ASD. Similar to the advent of CMA, the application of exome sequencing has enabled the study of SNVs and their role in disease. Recently, several high-throughput or next generation sequencing (NGS) studies of large cohorts have expanded the knowledge about the role of point mutations in ASD. Most of these studies have focused on *de novo* variation (66–70) and only a few have studied rare inherited variation (71–75) or combined both sources of information (74,76). Therefore, most studies have limited their research to variants of high penetrance that are expected to follow a dominant model of inheritance, as well as recessive conditions more common in inbred populations.

The first pioneering studies that applied exome sequencing to the study of ASD relied on a trio based design in order to identify *de novo* events (66–70). Although these studies vary in sample size, patient selection or variant calling and filtering parameters, similar conclusions can be drawn from them. First, the incidence of loss of function (LoF) or Likely Gene Disruption Mutations (LGD) variants seem to be approximately two times higher in probands than in unaffected siblings. Also, *de novo* LoF variants are more frequent in affected females than affected males. This is consistent with the prevalence differences between sexes and the hypothesis that females may need a higher genetic load to develop the disorder, a phenomena called the female protective effect. In addition, most studies confirm the direct relation between the incidence of *de novo* mutations and advanced paternal age. Overall, it seems that *de novo* LoF variants may contribute to up to 10% of ASD cases and they may affect hundreds of genes, highlighting the extreme genetic heterogeneity of the disorder. However, a few genes with recurrent LoF variants have emerged (Table 4) (66–70,77). Interestingly, an important proportion of genes with *de novo* mutations in ASD are also recurrently mutated in other neuropsychiatric disorders, such as epileptic encephalopathy, ID and schizophrenia (78).

Genes	# of events	FDR
CHD8	7	<0.0005
ARID1B, DYRK1A, SYNGAP1	5	<0.005
ADNP, ANK2, DSCAM, SCN2A	4	<0.01
CHD2, GRIN2B, KDM5B, POGZ, SUV420H1	3	<0.05
ANKRD11, ASXL3, ASH1L, BCL11A, CACNA2D3, CUL3, DIP2A, FOXP1, GIGYF1, ILF2, KATNAL2, KDM6B, MED13L, NCKAP1, PHF2, RANBP17, RIMS1, SPAST, TBR1, TCF7L2, TNRC6B, WAC, WDFY3, ZC3H4	2	<0.2

Table 4. Recurrent genes affected with *de novo* LoF events in several NGS sequencing studies with a trio based design. Adapted from (77).

These initial analyses have paved the way to subsequent studies targeting specific genes that have identified additional patients and reinforced their implication in ASD (79,80). Moreover, the detailed clinical characterization of patients sharing mutations in the same gene has enabled the identification of specific endophenotypes, that differ in their manifestations but result in ASD.

For example, mutations in *CDH8* result in similar phenotypic characteristics, such as macrocephaly, common dysmorphic features and gastrointestinal complaints, resulting in a definable syndromic subtype identified by a genotype first approach (81).

Regarding the identification of inherited causes of ASD by NGS, studies have followed two main approaches. The first approach has focused on genetic factors following a recessive model and have tried to identify biallelic genetic factors taking advantage of consanguineous families or families with distant shared ancestry. Related families have shown homozygous genetic variants in genes previously related to Mendelian disorders that result in attenuated phenotypes that lack some classical diagnostic features. In addition, studies in unrelated population show an increase of biallelic LoF variants in ASD patient compared to controls (71,73,75).

In the second and more recent approach, studies relying on larger sample sizes have combined both sources of evidence, inherited and *de novo* variants, in order to identify multiple hits in recurrently altered genes and pathways. Similar to previous reports, an excess of *de novo* LoF mutations in recurrent genes with a lower false discovery rate (FDR) was found among patients, especially among females. Unlike previous studies, an effect for *de novo* missense variants was also found. Inherited and *de novo* variants in recurrent genes were enriched in three main pathways that could play a role in the development of the disorder: chromatin remodeling, transcription and splicing and synaptic function (76). Interestingly, a second study found that private, inherited, truncated variants are enriched in probands compared to siblings and that their inheritance is biased towards a maternal origin. This is consistent with an oligogenic model, where transmitted mutations may increase the risk to develop the disorder and, again, consistent with a female protective effect (74).

1.4.5 Common SNVs

Apart from rare variants of high or moderate penetrance, common variants (present in >1% of population) of lower penetrance could also influence the risk of ASD. The contribution of common variants to complex disorders has been studied mainly through case-control studies interrogating Single Nucleotide Polymorphisms (SNP) throughout the whole genome, commonly

known as Genome-Wide Association studies (GWAs). To date, four large-scale GWAs studies that have been estimated to be able to detect variants of modest effect size have been carried out in ASD (82–87). These studies have found either single genetic loci associated to ASD that do not overlap between studies (5p14.1, 5p15.2, 20p12.1 and *CNTNAP2* in a secondary analysis) or no associations (82–85). In addition, studies targeting previously associated locus such as *MACROD2* in 20p12.1 have not replicated original findings (88). Moreover, a meta-analysis of three of the studies has shown that the combined data would result in a smaller association of each locus (89).

A possible explanation for these inconsistent results is that these studies may be underpowered and that larger cohorts are necessary to detect small effect signals. Some authors have pointed out that the low number of identified loci is similar to the results obtained in the first schizophrenia GWAs. Later on, these studies were proven to be underpowered given the promising results in recent studies with larger cohorts (87,90).

Even though no clear associated loci have emerged from GWAs, estimates of narrow sense heritability from studies of common variants in ASD have shown that they have a substantial contribution to risk (91,92). A first study estimating heritability found that common variants account for 60% of heritability in multiplex families and 40% in simplex families. This is consistent with an additive model in which each common variant exerts a small effect (91). A later study estimated narrow-sense heritability to be around 52%, mostly due to common variation. According to this study, rare variation has a small contribution to ASD variance, estimated to be around 3% (92).

To sum up, studies of common variation in ASD have resulted in few variants, each one contributing to a slight increment in risk. This explains only a small proportion of the familial aggregation observed in ASD. Taking into account rare variants, currently known genetic factors explain only a small proportion of the estimated heritability. Factors explaining this missing heritability include structural variants not genotyped by current methods, rare variants outside coding regions not yet detected, unaccounted gene-gene interactions and shared familial environmental conditions (93–95). The potential contribution of complex genetic rearrangements in regions of

segmental duplications (SD) and their inherent difficult study will be discussed in next sections.

1.4.6 Mosaic events

Another possible source of genetic variation contributing to ASD, although not inheritable, is mosaicism. Mosaicism is the presence of more than one genetically distinct cell line in a single organism. It is usually divided into somatic and germline or gonadal mosaicism, depending on the presence of different cell lines in the somatic tissue or the reproductive cells, respectively (96). Mosaicism has been observed in various Mendelian disorders, although its rate may be underestimated due to the difficulty of detecting low levels of mosaicism and the chose of the studied tissue, which is key to the identification of somatic mosaicism. Although brain would be the choice tissue for the study of mosaicism in ASD, more accessible tissues such as blood can provide a good approach. SNP array and NGS technologies are a good method to identify mosaic events at low frequencies (97,98), and have been recently applied to the study of mosaic mutations in the brain of ASD patients (99). The results of this study showed two cases with deleterious somatic point mutations, which may either be causative or contribute to the liability to ASD (99).

1.4.7 Inheritance models

As seen in the previous sections, a wide range of genetic factors of different effect sizes affecting hundreds of genes can influence the risk of ASD. ASD have a complex genetic architecture and the risk of developing the disease can be shaped by hundreds of variants of different magnitude of effect. As proposed for other complex diseases, the effect of a variant shows an inverse relationship with its frequency in the general population (93,100) (Figure 2). Genetic factors with a major effect usually have very low frequencies, since there is a strong selection pressure against their maintenance in the general population. In contrast, variants with small to modest effects have higher allelic frequencies in the general population and may also be present in control individuals.

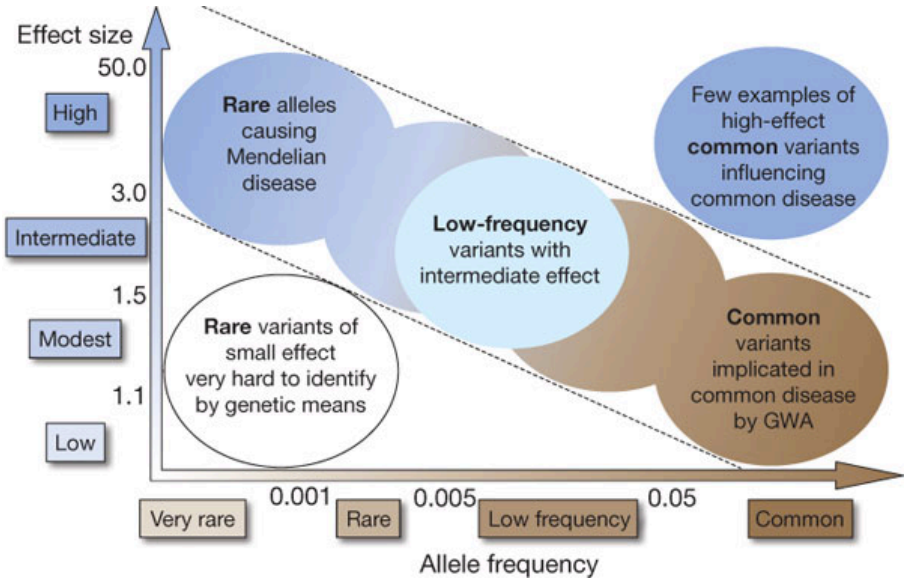


Figure 2. Relationship between allele frequency and size of effect. Different genetic factors with different effects contribute to complex diseases. Rare alleles causing Mendelian disease are identified by linkage studies or exome sequencing studies in selected families, whereas common variants with small effects are effectively detected by GWAs, provided that the study sample size is large enough. From (93).

In the case of ASD, different models of inheritance have been proposed to explain the interplay of genetic factors and their contribution to liability and variable expressivity. Each of them may contribute to part of the variability and may explain a subset of cases (87,101). The proposed models differ in their contribution of rare and common variants and range from Mendelian monogenic models to polygenic models, in which hundreds of variants of small effect shape the risk, as proposed in the common variant-common disorder hypothesis (87,102,103) (Figure 3).

In Mendelian models, a single genetic factor is sufficient to cause ASD. This single genetic factor may follow different inheritance patterns: autosomal dominant, autosomal recessive or X-linked, either dominant or recessive. Mutations in highly penetrant genes that cause syndromic or essential ASD follow this pattern of inheritance, such as *FMR1*, *MECP2*, *PTEN*, *CACNA1C*, *CNTNAP2*, *SHANK3*, *CHD8*, etc. In addition, cytogenetic defects, such as the 15q11-15q13 maternal duplication or deletions of 22q13 also follow this model. This model accounts for approximately 20-25% of cases found to carry a single and sufficient genetic factor. Interestingly, these rare highly penetrant variants are enriched in females and in individuals with lower IQ

scores, suggesting that females may need a higher genetic load to develop the disorder (102,103). However, additional genetic or environmental factors could also contribute to the disorder, since expressivity varies between individuals with the same genetic defect.

Although the Mendelian model may explain a fraction of the cases, there are evidences pointing towards a more complex genetic model, where various combinations of genetic and environmental factors may increase the risk of the disorder, but none is sufficient to cause it. Evidences supporting this model include studies observing that first-degree relatives of ASD patients often show subclinical traits associated with the disorder, commonly called "the broader phenotype". These include restrictive repetitive behaviors, deficits in social abilities or language delay (104–107). In addition, studies measuring autistic traits in the general population suggest a quantitative continuum and it has been hypothesized that the distribution of underlying genetic factors could be responsible for that gradient (108). In oligogenic models, relatives of affected individuals would carry a significant genetic load, consistent with the broader phenotype observed in families (87,102,103). Finally, genetic variants linked to ASD have been shown to influence social skills in unaffected people, supporting the notion that ASD lies along a continuum of traits in the population (109).

Proposed oligogenic models of inheritance take into account both rare and common variation and include various subtypes with different roles for rare and common events (87,102,103) (Figure 3).

Two of these models propose that ASD can arise from a combination of rare and common variants (Figure 3). In the first case, each parent would carry a few rare events of medium penetrance, whereas in the second, ASD would be caused by a combination of rare variants in a background of common risk alleles (87,103). *De novo* variants of smaller impact could also play a role and increase the risk of developing the disorder. These two models are supported by studies showing that private truncated mutations are enriched in probands compared to their siblings and that their inheritance is maternally biased (74). This is consistent again with a female protective effect, in which females would be carriers, but would need additional genetic variants to develop the disorder. In addition, the model is also supported by recurrent

CNVs inherited from an unaffected progenitor, such as deletions or duplications of 16p11.2 that show a reduced penetrance (60–64) (Table 3).

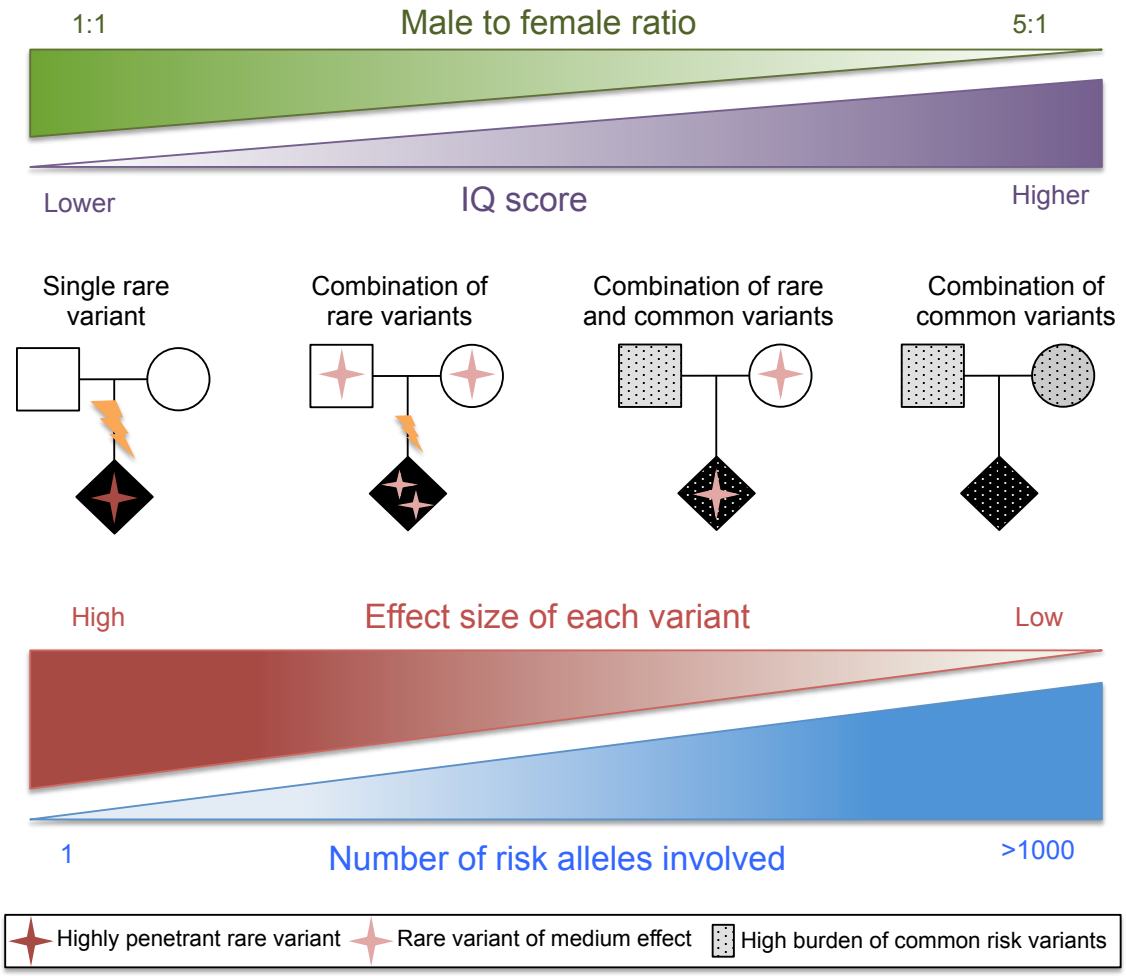


Figure 3. Schematic figure showing proposed models of inheritance in ASD. Mendelian cases (left) arise from highly penetrant mutations, that may follow different patterns of inheritance (autosomal dominant, recessive or X-linked). A *de novo* mutation is depicted on the figure. Oligogenic models include various combinations of rare and common variants (second left to right). Common variants have smaller effect sizes (red) and, therefore, a higher number of variants (blue) is needed to develop the disorder. Mendelian cases are enriched for individuals with lower IQ (depicted in purple) and females (depicted in green), as seen in studies comparing rates of *de novo* LoF mutations among groups.

Finally, a third polygenic model proposes that ASD arise from a genetic background of common risk alleles. In that case, the accumulation of a high load of low risk alleles would be sufficient to cause the disease, without contribution of rare alleles. These common variants are also present in the

general population, but accumulate in affected individuals and trespass the threshold of clinical expression. This is consistent with the *common disease-common variant* hypothesis, which refers to the theory that common diseases are due to genetic variants present in more than 1% of the population (110). This model may explain a subset of cases with higher IQ, where the male bias is even more marked and the rate of *de novo* LoF mutations is lower, and the previously explained fact that the same variants contributing to ASD are related to social skills in the general population. (102,109).

Ultimately, it should be noted that the role of common and rare inherited or *de novo* variants is not mutually exclusive and that both contribute to shape the risk and expressivity of ASD (87,102,103,111). In addition, environmental factors could also act as susceptibility or modifying or factors.

1.5 Environmental factors

Recent studies point towards an important contribution of environmental factors in ASD, with estimates of 50% of influence (45). This is supported by discordance in monozygotic twin pairs, which also suggests a role for environment (112). A few associations between environmental factors and risk of ASD have been described to date. However, it is difficult to reach solid conclusions due to the different methodology and lack of replication among studies. Until now, a few factors, such as maternal infection and drug exposure during pregnancy, as well as paternal age, are broadly accepted (113).

Regarding maternal infection, an epidemiological study in Sweden found out that maternal infection during pregnancy increased by 30% the risk of having a child with ASD (114). In addition, studies measuring maternal antibodies in pregnant women found that the presence of IgG reactivity to fetal brain was associated with having affected children (115). Animal models have also replicated this finding. By injecting isolated IgG from mothers with children with ASD (ASD-IgG) or typically developing children to a group of pregnant female rhesus monkeys, a subsequent study found that the offspring of mothers injected with ASD-IgG showed abnormal patterns of social interaction (116).

Paternal and maternal age have been also consistently associated with ASD (117,118). Although advanced paternal age is related to an increased rate of *de novo* mutations, other possible explanations for this phenomena are increased epimutations and accumulation of environmental toxins (113,117).

Other evidences for the influence of environment come from large environmental studies, such as the Childhood Autism Risks from Genetics and the Environment (CHARGE). This study found an association between various toxics contained in air pollution and ASD, which is also supported by previous associations in smaller studies (119,120). For example, exposure to nitrogen dioxide and other traffic-related pollutants during the late period of pregnancy and early childhood has been associated with an increased risk of ASD (119,120). Studies focusing on perinatal and postnatal conditions have also identified several factors that could increase the risk. Some of these are: abnormal fetal presentation, umbilical cord complications, fetal distress, multiple birth, small size of gestational age and low birth weight (113).

Environmental factors can act through epigenetic mechanisms, which provide a link between genetics and environment. Epigenetics comprises the study of changes in gene function that are mitotically and/or meiotically heritable and that do not entail a change in DNA sequence (121). Complex traits such as height, fertility or food metabolism are environmentally influenced and its modifications can be transmitted by epigenetic mechanisms (122). The study of epigenetics in ASD is relatively new, but there is evidence that methylation changes, as well as chromatin remodelers, could be implicated in the disorder (123). For instance, studies have found genome-wide hypomethylation rates in patients but not in matched siblings or controls (124). In addition, regions that have been associated to ASD have been found to be located near imprinted sites. Finally, numerous chromatin remodelers are mutated in ASD, as well as in other neurodevelopmental disorders (125).

1.6 System biology approaches to ASD

In the recent years, new genetic approaches have uncovered hundreds of genetic variants implicated in ASD. However, the effects of the deregulation of these genes on molecular pathways, the functioning of the cell and, ultimately, on neural circuits and behavior are not so well known. System

biology tools can help understand the mechanistic effects of the altered genes by integrating different sources of information, such as genetic variants identified by GWAs or NGS studies, expression data or physical interaction networks of proteins (126).

Regarding expression, transcriptional studies can help to identify and understand deregulated gene networks in ASD. Although, this approach is limited by the difficulty to obtain neural tissue, several studies have analyzed expression changes in brains of ASD patients. The first studies used a case-control approach to measure differential expression by microarray technology and came up with few reproducible results. Overall, they identified alterations in neuronal systems, implicating glutamatergic neurons, and alterations in immune related genes, but few genes were coincident among studies (127–130). More recently, two studies using weighted correlation network analysis (WGCNA) found similar altered modules, one related to neuronal and synaptic functions and a second module enriched for microglia immune genes. Interestingly, the neuronal module was enriched for genes previously identified by GWAs studies, suggesting a causal role in the disorder (131,132).

Mapping altered genes to expression networks is also a good strategy to understand which mechanisms are affected by genetic factors. Some authors have suggested the importance of mapping genes to developmental expression networks (133), since ASD are classified as a neurodevelopmental disorder and the alterations leading to it begin during prenatal time (126). The first study to do that mapped ASD candidate genes and GWAs associated loci to co-expression modules based on gene expression in the human brain (134). Two neuronal modules were enriched in genes harboring both common and rare variation. In addition, its expression was found to vary among developmental stages, showing the importance of neural development in ASD (134).

Later on, subsequent studies made use of the availability of NGS studies of ASD and ID and transcriptome data from fetal and adult brain tissues. To this end, big-data initiatives such as the Genotype-Tissue Expression (GTEx) project or the Allen Brain Atlas provide key information to integrate genotypes and expression across a whole range of tissues, specific brain areas and

precise time-points. Making use of this important amount of data, a study identified that ASD genes were enriched in modules, such as synaptic functions and translational regulation, expressed during cortical development. Moreover, the study found that ASD and ID candidate genes presented distinct patterns of expression (135). A second study mapped ASD strong candidate genes derived from sequencing studies to very specific time points and locations: the midfetal layer 5/6 cortical projection neurons (136).

Another approach to identify common deregulated pathways in ASD is to make use of physical interaction data, either from experimental approaches or from curated literature, such as protein-protein interaction (PPI) networks. This approach was first applied to results from exome sequencing studies studying the contribution of *de novo* variation and found that genes carrying LoF mutations conformed a highly interconnected network, related to synaptic transmission and regulation of translation (66,67). Other studies taking into account both inherited and *de novo* variation also found a tight network, with four clusters including previously related pathways such as synaptic transmission and transcriptional regulation (Figure 4) (76). A posterior study constructed PPI networks to look for interacting partners among 35 ASD candidate genes and found novel interactions and unexpected connections between syndromic ASD genes (137). Interestingly, another study analyzing interactions between spliced variants of brain expressed ASD risk genes identified new interacting partners, showing the importance of taking into account isoform networks (138).

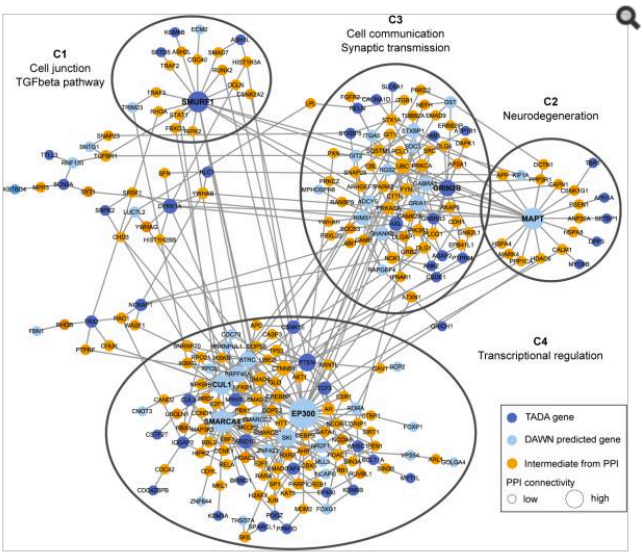


Figure 4. PPI network of candidate genes identified by integration of data from exome sequencing studies, showing four clusters, including previously related networks of synaptic transmission and transcriptional regulators with black circles. From (76).

Finally, a recent approach to integrate genes in network modules reflecting underlying phenotypes, called Network-Based Analysis of Genes (NETBAG), has been developed and applied to ASD (57). Genes altered by *de novo* CNVs in ASD cases were mapped into the network and were found to be primarily related to synapse development, axon targeting and neuron motility. A more recent study applied the same method to integrate genes affected by *de novo* truncating mutations from several exome sequencing studies and again found that cortical circuits seem to be particularly affected (139). Interestingly, the study also found that genes affected by truncating SNVs in affected females had higher expression in brain compared to genes with truncating mutations in males (139). This shows that system biology tools not only can provide insights in the pathophysiology of ASD, but also into its genetic architecture. A summary of the clinical characteristics, genetic causes and altered mechanisms in the first section are summarized in Figure 5.

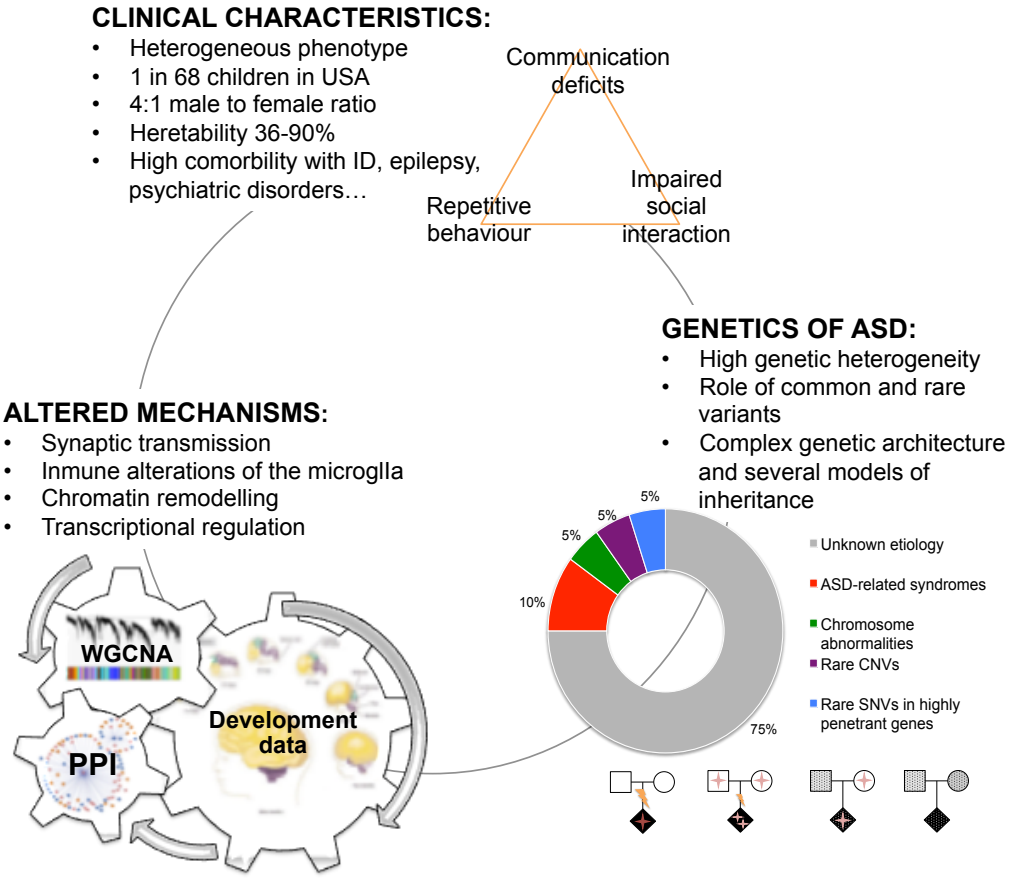


Figure 5. Summary of the main clinical characteristics, genetic causes and altered mechanisms in ASD.

1.7 Evaluation and medical management of ASD

ASD are usually diagnosed in the early childhood and can be reliably diagnosed by the age of two (140). Early diagnosis is important, since early intervention can improve language and communication skills and adaptive behavior in children (141,142).

Guidelines for ASD diagnosis state that it should involve a two-stage process. The first stage comprises surveillance and developmental screening through regular medical checkups with a pediatrician or a primary care physician. If concerns about the child's development are raised, children should be referred for additional evaluation. In this second stage, thorough evaluation by a team of health professionals from a wide range of specialties should be performed (143,144).

Once the diagnosis is established, a management program should be established. Follow-up should consider educational therapies and interventions in order to help mitigate the core features of ASD and treatment of possible associated medical problems, such as sleep dysfunction, psychiatric conditions or seizures (145).

1.7.1 Surveillance and screening

Surveillance has been defined as the ongoing process of identifying children who may be at risk of developmental delays, whereas screening is the use of standardized tools at specific intervals to support and refine the risk (146). According to current guidelines, routine developmental surveillance and screening for ASD should be performed on all children to identify those at risk for any neurodevelopmental disorder and, specifically those at risk for ASD (143,144). The American Academy of Pediatrics (APP) has even recommend that all children between 18 and 24 months should be screened for ASD during regular medical visits with their pediatrician (143).

Surveillance should begin at regular care visits with the children pediatrician or primary care physician. It should include a careful family history to identify family members diagnosed with neurodevelopmental disorders. Professionals should pay especial attention to siblings' developmental history, since the risk of ASD in siblings of children with ASD is increased ten times compared to the general population. Parents' concerns about their child development and

behavior should also be taken into account, since several studies have shown that they provide useful information. Moreover, standardized parental reports have been shown to be effective for early screening (147–149). In addition, the health professional should also ask for specific questions regarding achievement of developmental milestones, interact and explore the child, to check if the elicited response corresponds to that of a typically developing child (143,144).

Standardized screening tools should be administered when a specific concern about ASD has been raised as a result of surveillance protocols. Specific tools for the detection of ASD have been developed, although none has been validated for children younger than 18 months. Some of this, such as the Checklist for Autism in Toddlers (CHAT) (150), Modified Checklist for Autism in Toddlers (M-CHAT) (151) and the Pervasive Developmental Disorders Screening Test-II Primary Care Screener (PDDST-II) (152) were specifically designed for early detection.

1.7.2 Diagnostic evaluation

Once a child has been identified through specific screening, he or she should be referred without delay to additional services for a comprehensive evaluation and early intervention (143).

I. Clinical assessment

A detailed clinical assessment is an important part of the diagnostic evaluation, since it can identify underlying associated medical conditions. A complete medical history, including detailed prenatal, perinatal and postnatal development and family history should be obtained. Patients should also undergo clinical examination. Clinicians should pay attention to dysmorphic features, head circumference, neurological abnormalities and skin spots in order to identify specific syndromes comorbid with ASD. Audiological evaluation should also be performed in all children, especially those with language delay (153).

As part of the diagnostic evaluation, genetic testing to identify a genetic etiology should be performed. There are several genetic tests available for the diagnosis of ASD (153).

X-fragile testing

Approximately 0,5% of children with ASD have X-fragile syndrome and about 20% of children with X-fragile syndrome present ASD features. Due to this overlap, the American College of Medical Genetics and Genomics (ACMG) recommended X-fragile testing for all boys being evaluated for ASD. Regarding girls presenting with ASD, the ACMG does not recommend X-fragile testing, unless there are suggestive items such as positive family history, a suggestive phenotype or associated features (premature ovarian insufficiency, ataxia or tremors) in relatives (154).

Karyotype

G-banded karyotype has been the traditional approach to diagnose patients with a neurodevelopmental disorder of unknown origin. The diagnostic yield of karyotype is estimated to be around 3% excluding chromosomal syndromes clinically recognizable, such as Down syndrome. However, since the advent of CMA in 2004 (51,52), traditional karyotype has been replaced (155).

Chromosomal microarray

Since around 15% of individuals with ASD carry a CNV detectable by CMA, its diagnostic yield in ASD is much higher than that of previous techniques such as karyotype or X-fragile testing (48,53–56). Because of this, the ACMG and other authors have proposed chromosomal microarray as the first diagnostic tool for the diagnosis of ASD, developmental disabilities and congenital abnormalities (154,155).

However, the adoption of CMA as a first tier tool also poses challenges, such as the identification of Variants of Unknown Significance (VOUS). This challenge can be addressed by using a suitable level of resolution, which should offer enough sensitivity to identify all known recurrent and nonrecurrent genomic rearrangements, but minimize unexpected findings (153). In addition, the development of databases, such as the International Standard Cytogenomic Array (ISCA) Consortium Database (www.iscaconsortium.org), DECIPHER (www.decipher.sanger.ac.uk) or the DataBase of Genomic Variants (DGV, www.projects.tcag.ca/variation/) is helping in the classification of CMA findings, by further defining variability in control population and recurrent events in affected patients. Finally, genetic

study and detailed phenotypic characterization of the progenitors can also help to interpret VOUS (153).

Single-gene testing

Several single-gene disorders, apart form X-fragile, have been associated to ASD. For example, deletions or mutations in *MECP2* are associated to Rett syndrome in girls and duplications and mutations with a milder effect have been found in boys with ASD and severe ID. In addition, mutations in *PTEN* have been found in approximately 5% of patients with ASD and in a higher frequency among patients with macrocephaly (154). With the advent of NGS sequencing, it is now possible to sequence multiple genes, all protein-coding parts of all genes or the entire genome in a single test, relegating single-gene testing to confirm clear clinical suspicions (154,156).

Exome sequencing

Due to the extreme genetic heterogeneity of ASD, exome sequencing and other NGS based methods have been proposed as diagnostic tools for ASD (154,156). However, few studies have evaluated the diagnostic yield of this approach in ASD or other neurodevelopmental disorders (157,158). In a recent study where both CMA and exome sequencing were applied to a group of patients with ASD, the diagnostic yield of the two methods was estimated to be 9.3% and 8.4% respectively (158). This diagnostic yield was higher in patients with syndromic ASD (158). Although studies in larger cohorts are needed, these data show the utility of exome sequencing in ASD. Moreover, exome sequencing can also detect CNVs, which is expected to increase the diagnostic yield an additional 10% (159), with a total diagnostic yield combining the detection of SNVs and CNVs of approximately 20%. However, studies comparing the detection rate of CMA and exome sequencing are needed before the definition of the optimal first-tier genetic test.

Similar to the application of CMA, the clinical use of exome sequencing poses important challenges, such as the finding of VOUS and the increased probability of incidental findings not related to the phenotype. This supposes a challenge for genetic counseling (160–163). Several recommendations have been proposed for the report of incidental findings and the contents of informed consent (164–166). Finally, the application of whole genome

sequencing (WGS), now restricted to research, to the clinics is also expected to increase the diagnostic yield in ASD. Since the potential of incidental findings and VOUS will increase substantially, the study of genetic variation in non-coding regions in the general population will help to interpret the results (72,167).

Based on current recommendations and recent advances in the molecular diagnosis of ASD, a model for genetic testing is proposed on Figure 6.

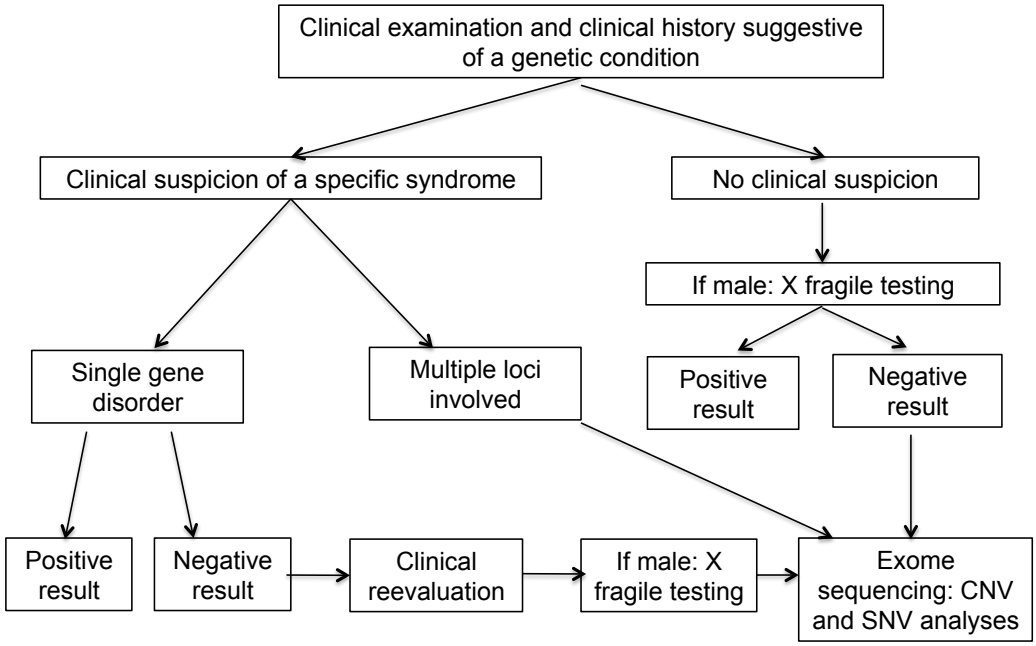


Figure 6. Proposed tiered genetic evaluation for ASD. Exome sequencing is proposed as the first tier diagnostic tool for ASD patients without clinical suspicion of a known syndrome. The proposal relies on the higher expected diagnostic yield of exome sequencing, by the combined detection of CNVs and SNVs. Since X-fragile is caused by a triplet expansion and therefore cannot be detected by exome sequencing, testing is recommended to all males, especially if there is a suggestive family history of X-fragile related disorders. Due to its cost effectiveness, exome sequencing is also the choice diagnostic approach for patients with a clinical presentation suggestive of a known syndrome with genetic heterogeneity. Single gene disorder testing is indicated only in patients with a clinical suspicion of a monogenic syndrome.

As well as genetic testing, additional tests should be considered when evaluating a child with ASD. Due to the high incidence of seizures among children with ASD, electroencephalogram (EEG) is indicated, especially in children with clinical signs of seizures or those with language regression. Neuroimaging studies, such as magnetic resonance imaging (MRI), are also indicated in children with ASD and additional neurological signs in order to

detect brain malformations. Although there is not a specific biomarker for ASD, metabolic studies are indicated in patients with ASD and additional suggestive clinical signs, such as organomegaly, coarse features, etc. (143,168).

II. Psychological assessment

Given that there are not biological biomarkers that confirm the diagnosis of ASD, psychological and behavioral assessments are the most common instruments for the diagnosis. There are several tools, but the two most widely used and with the highest sensitivity and specificity are the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS) (169). The ADI-R is a semi-structured interview conducted to the parents by a highly trained clinician. It focuses on three main areas, as described in the ICD-10 and DSM-IV: reciprocal social interaction, communication and language and restricted repetitive stereotyped interests and behaviors. It is suitable for children from 18 months (170). The ADOS is a semi-structured assessment of behavior in which the clinician scores behaviors categorized in three main areas: communication, social interaction and play. It has four modules, each of one specific for different developmental and language stage (171). Some authors have suggested that the combined use of both tools reaches higher levels of accuracy and recommend their combined use (169).

1.7.3 Management

I. Medical management

Currently, there are no curative treatments for ASD. Therefore, medical management, psychopharmacological treatments or educational interventions aim to reduce the symptoms and improve the quality of life of affected individuals (172).

Regarding medical attention, children with ASD have specific medical needs in addition to the basic healthcare that all children must receive. As commented before, seizures, gastrointestinal problems and sleep disturbances are frequent in children with ASD. Children with underlying conditions, such as X-Fragile or tuberous sclerosis, may present associated medical concerns. Morbidity and mortality are increased among individuals

with ASD, and the data suggests that mortality is 2.4 to 2.6 times higher compared to that of the general population (131,152).

Regarding seizures, anticonvulsant treatment should be used depending on the specific seizure type and based in regular criteria. Abnormalities in EEG are common in children with ASD. However, since the consequences of abnormal EEG without frank seizures on development are unknown, treatment is not recommended.

Even though gastrointestinal problems are more frequent among children with ASD, specific testing is not recommended unless clinical signs are present. Treatment of gastrointestinal complaints is important because they may lead to sleep disturbances or aggression outbursts.

Sleep disturbances, they can have a great effect on functioning and quality of life of the children and family. Behavioral therapies are the first line of treatment for those cases with no identified cause. Pharmacological management is also possible, although few studies have been conducted (131,152).

II. Educational interventions

Educational interventions in children with ASD aim to improve socialization increase adaptative behaviors and minimize interfering behaviors. Therapies may help to improve the core symptoms of ASD, although they usually remain over life (145,173).

There is a wide range of educational interventions, which are usually divided into three types: behavior analytic, developmental or structured teaching. Although these approaches may vary in their strategies and there is not a global consensus (174), experts agree that effective interventions should include early intensive intervention, a low student-to-teacher ratio and family involvement. In addition, they should provide opportunities for interacting with typically developing peers and include a curriculum that considers acquisition of communication, social and adaptive skills as well as cognitive and academic skills (145,173).

The most common methodologies include Applied Behavior Analysis (ABA), structured teaching (also called Treatment and Education of Autistic and

Related Communication Handicapped Children or TEACCH), specific language therapies such as the Picture Exchange Communication System (PECS) and occupational and sensory integration therapies (145,173).

Apart from early intervention therapies, programs should also include transition for older children and adolescents. These programs should consolidate and expand social, communication and behavior skills in order to achieve independence. Finally, transition to adolescence and adulthood should include an individualized plan for employment according to each individual's needs and abilities (145,173).

III. Psychopharmacology interventions

Pharmacological treatment should be considered in individuals presenting disruptive behaviors that cause impairment in daily activities, such as self-injury, compulsions and stereotypies, sleep disturbances, irritability or anxiety. An important proportion of individuals with ASD are treated with diverse medication (around 45% of children) and this percentage is increased in older individuals (75% of adults) and those with lower adaptive social skills (145).

The most commonly prescribed drugs are selective serotonin-reuptake inhibitors (SSRIs), atypical antipsychotic agents, stimulants and α2-adrenergic agonist antihypertensive agents. Even though SSRIs have several adverse effects, several studies have shown their efficacy in targeting core symptoms, including repetitive behaviors, irritability or aggression. Risperdone, which is an atypical antipsychotic agent, is among the most used drugs for the treatment of irritability in children and adolescents with ASD. Stimulants are also effective for the treatment of hyperactivity, impulsivity and inattention in children with ASD. Finally, α2-adrenergic agonist antihypertensive agents can help reducing irritability, impulsivity and repetitive behaviors (145).

IV. Family support

Management should also consider family support. Several studies have shown that stress and depression are more frequent among parents of children with ASD than among parents of typically developing children or of children with other neurodevelopmental disorders (175,176). In order to support parents' needs, health care professionals should provide them with

education about ASD, access to resources, emotional support and guidance in therapy and educational decisions. Genetic counseling should be included as part of the family support and should be offered to parents, as well as to other family members interested (145).

The main steps and points to take into consideration regarding evaluation and medical management of patients with ASD are depicted in Figure 7.

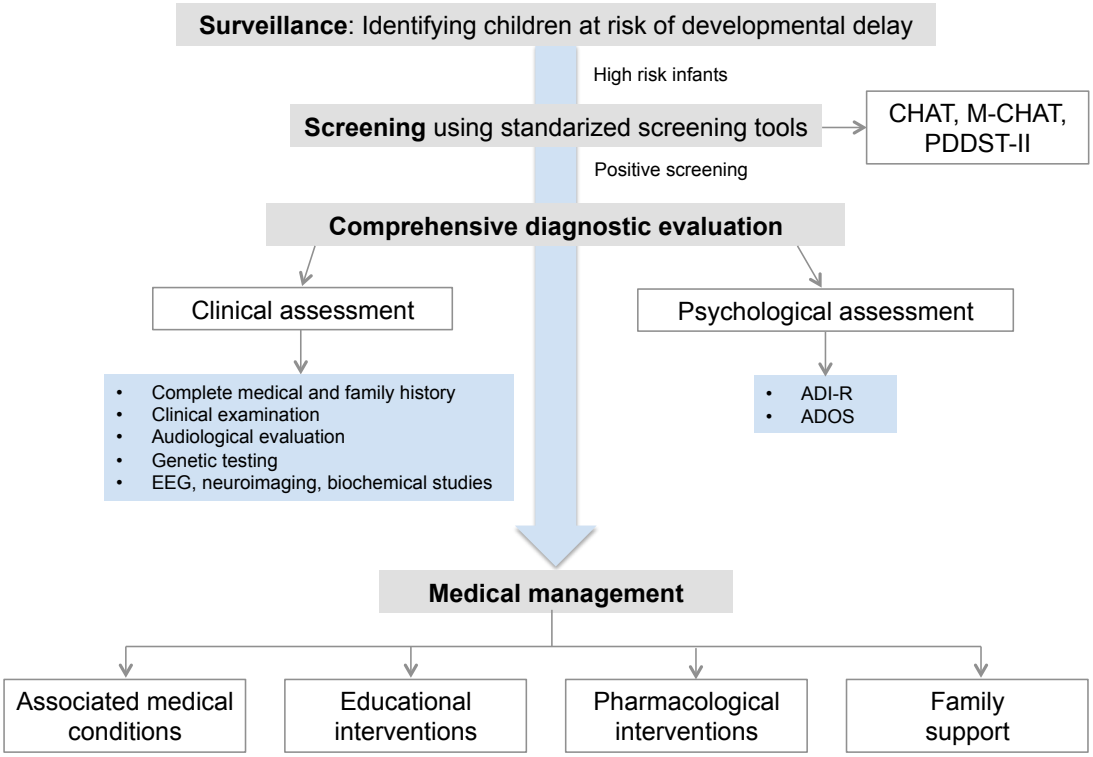


Figure 7. Schematic representation of the main steps involved in the evaluation and medical management of patients with ASD.

2. Segmental Duplications

Several authors have suggested that complex rearrangements, occurring mostly in polymorphic multi-copy number variants located in SD, could explain part of the missing heritability in complex disorders (94,177,178). SD are defined as segments of DNA that range from 1 to 400kb, are present in at least two locations in the genome and share a high degree of identity (90-100%). They are also known as low-copy repeats (LCR) and are estimated to account for a substantial part of the human genome, approximately 5% (179,180). They can predispose to various types of chromosomal rearrangements, such as duplications, deletions, inversions, translocations and marker chromosomes through misalignment of non-allelic homologous regions and unequal crossing over.

The study of segmental duplications is of great interest, since they have a crucial role in human and primate evolution and in the predisposition to recurrent genomic disorders. More than 25 genomic disorders are thought to be mediated by non allelic homologous recombination (NAHR) between LCR (179–184). Intriguingly, some of the duplicated regions have arisen in the primate lineage and are even specific of the human lineage (185). For instance, SD flanking the 16p11.2 region, which has been extensively associated to ASD, have been found to be duplicated only in the modern human lineage. This recent rearrangement predisposes to NAHR and, therefore, to deletions and duplications associated to ASD (186).

This example raises intriguing questions about the relation between genes contained in SD, human specific traits and neurodevelopmental disorders. Given that an important part of SD have been generated in the primate or human lineage, they are expected to be enriched in genes with a role in cognition and behavior. These genes could play a role in ASD and contribute to its missing heritability. Although SD could represent an important part of genetic variability, these regions have been largely unexplored due to its inherent complexity. Several techniques have been applied to identify and refine the structure of LCR, such as array comparative genomics (aCGH) or SNP microarrays. The advent of NGS offers new opportunities to further define duplicated regions and breakpoints in recurrent rearrangements.

2.1 Type of segmental duplications and distribution in the human genome

SD can be classified into two groups: interchromosomal and intrachromosomal (179,180). Interchrosmosomal SD localize in non-homologous chromosomes, whereas intrachromosmal SD are located in the same chromosome and, usually, in the same band. Both groups of SD are enriched in pericentromeric and subtelomeric regions, although intrachromosomal SD are usually more distant to centromeres or telomeres (187). The distribution of SD in human chromosomes is not uniform, and some chromosomes harbor more LCR than others (Figure 8). Since SD are a source of genetic variability, CNVs and copy neutral structural variants tend to localize near these regions (188).

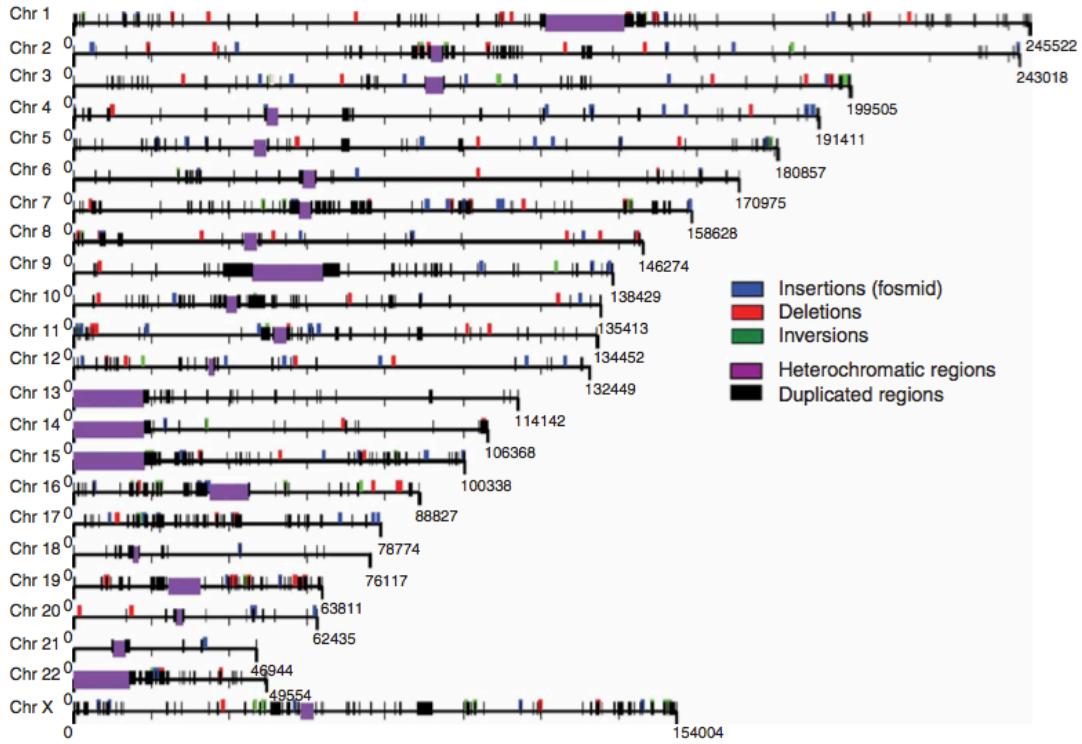


Figure 8. Map of the human genome showing the distribution of SD (in black) across all chromosomes. SD are enriched in subtelomeric and pericentromeric regions and structural variants (insertions, deletions and inversions) in regions surrounding SD. Adapted from (188).

2.2 Origin and role of segmental duplications in evolution

SD have evolved in a rapid manner through the evolution of primates, especially during the divergence of great apes and humans (178).

Chromosomal rearrangements have been particularly important in centromeres and telomeres, which are thought to be more tolerant to the addition of extra genetic material. This process has occurred during different time points, although possibly not at a constant rate (180,189).

SD have an important role in the apparition of new genes and, therefore, new functions (178). In most cases, gene duplications would result in non-functional pseudogenes, termed paralogous copies. Paralogous copies may be complete or may lack necessary elements for transcription, such as regulatory elements or 5' exons. In addition, they tend to accumulate variants, that may result in premature stop codons and truncated proteins. In a small proportion of cases, gene duplications may result in the apparition of new functions due to positive selection. In that case, a new copy of the gene may evolve without functional constraints, since its progenitor copy would still retain its original role. An example of the acquisition of new functions is the duplication of the X-linked opsin genes, which enabled the evolution of trichromatic color vision (180,189).

2.3 Segmental duplications and disease

Genomic disorders were first described as conditions that result from rearrangements of genetic material due to the recombination of highly homologous stretches of DNA (190). These rearrangements usually result in deletion or duplication of dosage sensitive gens in the single copy region flanked by SD.

Chromosomal rearrangements between SD give rise to contiguous gene syndromes, as well as monogenic traits, depending on the distance between SD and the number of genes included in the flanked segment. Among monogenic traits, some of the most common are adrenal hyperplasia due to 21-OH deficiency, spinal muscular atrophy, familial juvenile nephronophthisis 1, Charcot–Marie–Tooth (*CMT1A*). Examples of contiguous gene syndromes include Williams-Beuren syndrome (WBS), Prader-Willi and Angelman syndrome, Smith-Magenis syndrome or DiGeorge syndrome (46).

2.4 Rearrangement mechanisms

There are three main mechanisms that can give rise to genomic rearrangements: NAHR, non-homologous end joining (NHEJ) and Fork

Stalling and Template Switching (FoSTeS). Of those, only NAHR is mediated by SD, being associated with a higher risk of occurrence and recurrence of genomic disorders (191).

2.4.1 Non allelic homologous recombination

NAHR is responsible of most recurrent rearrangements and can give rise to deletions, duplications and inversions. It occurs due to misalignment of non-allelic LCR and subsequent unequal crossing over. It occurs mostly in meiosis, although it may also happen in mitosis. If duplicated sequences are found in the same chromosome and are in direct orientation, NAHR can give rise to either deletions or duplications. If duplicated sequences are found in the same chromosome but in inverse orientation, NAHR can result in inversion of the in-between fragment (Figure 9).

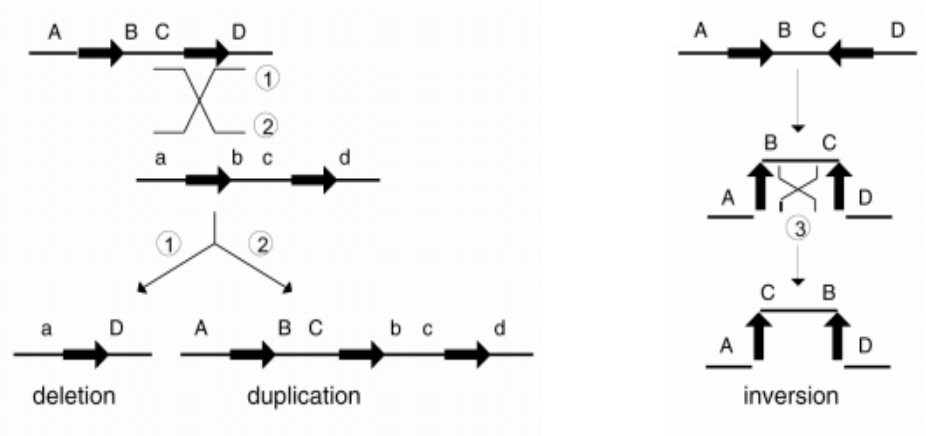


Figure 9. Schematic representation of NAHR. A) Repeats in the same orientation can generate duplications or deletions. B. Blocks in the opposite direction can result in an inversion. Adapted from (191).

During meiosis, NAHR can occur between SD on the same chromatid (intrachromatid), on sister chromatids (interchromatid) or on the homologous chromosome (interchromosomal). If it takes place between two segments in direct orientation, interchromatid and interchromosomal NAHR can give rise to reciprocal deletions and duplications. However, intrachromatid NAHR can only lead to a deletion and circular DNA molecule without centromere that that will not segregate during cell division (Figure 10) (192). This explains why the frequency of deletions in reciprocal genomic disorders is usually

higher than that of duplications, since one of the mechanisms, intrachromatid NAHR, only leads to deletions.

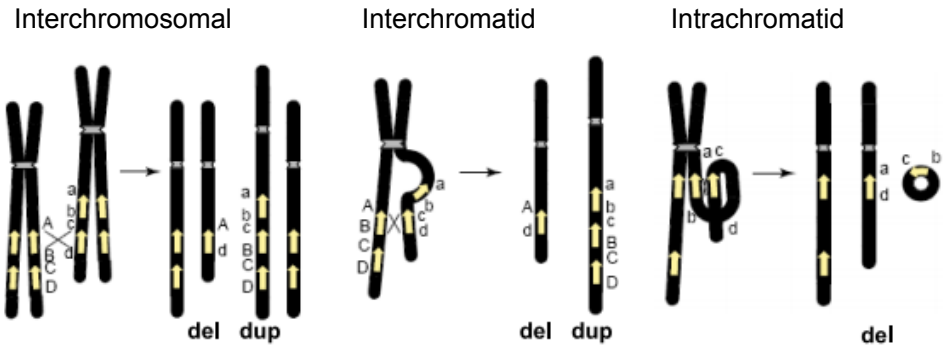


Figure 10. Mechanisms for non-allelic homologous recombination (NAHR). The three classes of NAHR are interchromosomal, interchromatid and intrachromatid. Interchromatic or interchromosomal can result in a deletion or duplication, whereas intrachromatid NAHR can generate a deletion and a circular DNA molecule without a centromere. Adapted from (191).

2.5 Methods to study multi-copy variants in segmental duplications

Despite the fact that genes inside SD play a major role in evolution and disease, these regions remain largely unstudied. Conventional techniques to detect CNVs, such as aCGH or SNP microarrays, have a limited detection rate in multi-copy number variants sites due to the high similarity of these regions. Their power decreases as the underlying number of copies increases. aCGH has been successfully applied to the genome-wide identification of CNVs, but the identification and fine characterization of variation in SD has proven more difficult. Due to probe cross-hybridization, the specific duplicated or deleted copy can be quite difficult to define(182,193,194). SNP microarrays are neither a good technique for studying multi-copy regions, since probes covering SD are underrepresented in the current genome-wide platforms. Although paralogous sequence variants (PSVs), which are positions differing between duplicated copies, can be genotyped with with SNP microarrays and an important proportion of SNPs are in fact PSVs, cross-hybridization can also confound the results (195–197).

However, the advent of NGS provides new tools to study SDs, with information about the number of copies (by read-depth) and the identity of duplicated sequences (by genotyping PSVs). This approach has been successfully used in WGS data of control individuals, providing a reference map of SDs and PSVs to be used in further studies (198,199).

3. The 7q11.23 region: a dosage-dependent effect on neurodevelopment

One of the complex regions in the genome is the 7q11.23 band. Despite its complex genomic architecture, it has been characterized in detail (200–202). It contains three different blocks of SD that predispose to various rearrangements: reciprocal deletions and duplications, inversions and polymorphic CNVs. Interestingly, the reciprocal deletion and duplication of the single copy genes flanked by SD result in mirror phenotypes, in terms of behavior and other clinical conditions. The deletion causes WBS (OMIM #194050), which is associated to mild to moderate ID, overfriendly personality and relative language preservation. On the other hand, the reciprocal duplication (OMIM #609757) results in speech delays and alterations and deficits in social interaction (203). The study of this region provides an opportunity to understand the involvement of the dosage sensitive genes in language development and behavior, as well as the effect of modifying factors, such as alleles with varying degrees of activities or genes outside the affected region, in the phenotype.

3.1 7q11.23 genomic architecture

The singly copy region of the WBS locus is flanked by three large SD, which share a high degree of homology (≈98-99%) located in the centromeric (c), medial (m) and telomeric (t) sides. Each SD is composed of three blocks with different order and orientation, named A, B and C. The centromeric and medial blocks are in the same orientation, but different order, whereas the telomeric blocks are ordered as the centromeric, but in inverse orientation (Figure 11) (200).

Regarding gene content, the blocks contain multiple transcriptionally active units, although some are transcriptionally active. The B block contains *GTF2I* and *GTF2IRD2*. *GTF2I* is considered the main candidates for the neurocognitive profile and *GTF2IRD2* a possible modulator (Figure 11) (204–

207). Of the three copies of *GTF2I*, only one (located in the medial B block) is functional. The rest of the copies (located in the centromeric and telomeric B blocks) are transcribed but result in a truncated protein, since they lack the 5' region of the gene (208). Regarding *GTF2IRD2*, there are two functional copies of the gene located in the medial and telomeric blocks, while the centromeric copy is not transcribed since it lacks the first two exons (188).

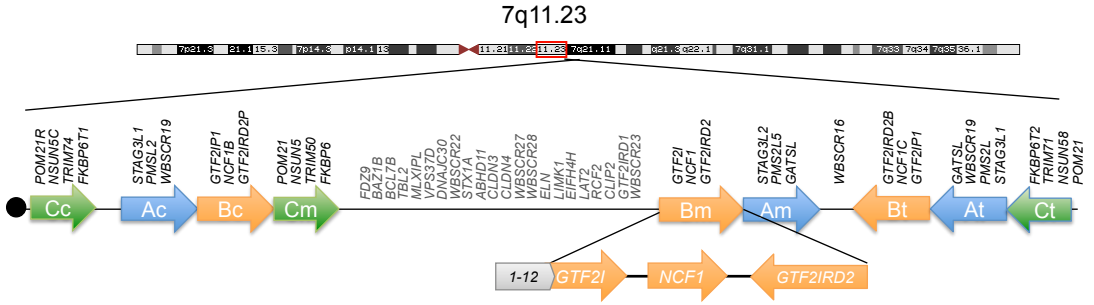


Figure 11. Schematic representation of the 7q11.23 region. The centromeric (c), telomeric (t) and medial (m) SD are showed in arrows with their respective blocks colored in blue (A), orange (B) and green (C). A magnified image of centromeric block B shows the functional copies of *GTF2I,* with the first twelve exons present only in the single copy region in grey.

3.1.1 Deletion

There are two types of recurrent deletions that result in WBS syndrome. The most frequent is a 1.55Mb deletion that occurs in 87% of cases as a result of NAHR between the medial and centromeric B blocks. Around 10% of patients present a larger deletion of 1.8Mb due to a crossing over between centromeric and medial A blocks (Figure 12). Atypical deletions mediated by other mechanisms can be found in approximately 3% of patients (209,210).

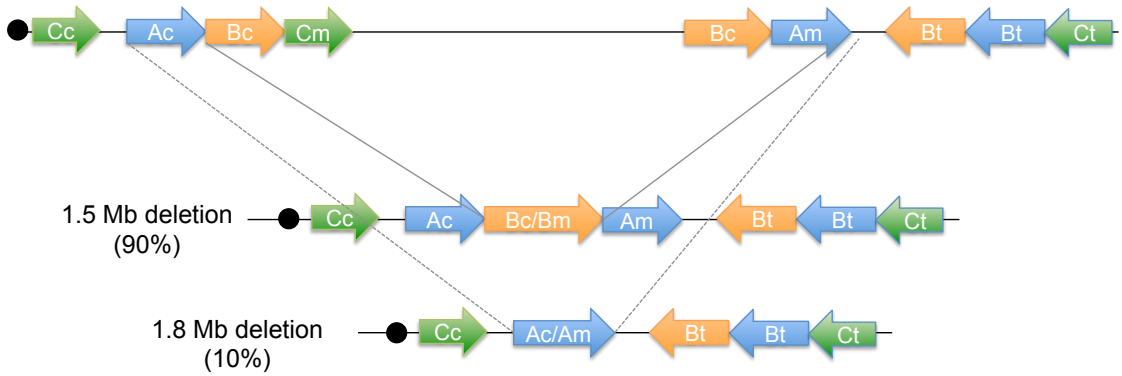


Figure 12. Schematic representation of the two most frequent deletions in WBS

syndrome. The size of the deletion depends on the type of blocks where the unequal crossing over occurs.

It has been hypothesized that deletions between B block may occur more frequently than between A blocks due to the their higher degree of homology (99.6%) and because the genomic distance between B blocks is shorter than between A blocks (209). Haplotype analysis have shown that most of the deletions, approximately two thirds, arise from intechromosomal recombination rather than from intrachromosomal crossing over (either inter or intra-chromatid) (209,211,212). No bias depending on paternal origin of the transmitting disease chromosome has been reported (209).

3.1.2 Duplication and triplication

Reciprocal duplications of the WBS region should occur at the same frequency of deletions due to interchromosomal crossing over. However, the first case was not described until 2005 (213). Since then, and thanks to the application of CMA, many additional cases have been reported. The clinical characterization of new cases has permitted to delineate a novel syndrome, estimated to affect 1:13,000–1:20,000 individuals However, this frequency may be underestimated due to the variable and milder phenotypic expression of this syndrome compared to WBS. The milder phenotype explains that the duplication is inherited from an affected parent in approximately 27% of probands (215).

In 2010, a patient with a triplication of the region was described. The triplication was detected by chromosomal microarray and confirmed by Multiple Ligation Probe Amplification (MLPA) and Fluorescence In Situ Hybridization (FISH). Since the triplicated region did not include the common deleted genes *FZD9* and *FKBP6,* the triplication does not seem to be caused by NAHR. The patient presented with ID, a severe expressive language delay, behavioral problems and dysmorphic features similar to those found in individuals with the duplication. These findings support the idea, that the WBS region is sensitive to dosage changes (216).

3.1.3 Other rearrangements

I. The paracentric inversion of 7q11.23

A polymorphic inversion has been identified in 7q11.23 occurring in 25% to 33% of WBS transmitting progenitors (217) and in a similar frequency in transmitting parents of children with a *de novo* reciprocal duplication (unpublished data). In the general population, however, it is found in a far lower percentage, around 5.8% of individuals (209,215,218). Although carriers of the inversion have no abnormal phenotype, it predisposes to rearrangements of the region by facilitating unequal crossing over during meiosis. Therefore, individuals carrying the inversion have an increased probability of having a child with a duplication or deletion of the region compared to the general population. The recurrence risk for carriers of the inversion is 1 in 1750 compared to 1 in 9500 for non carriers (218).

The inversion is most likely due to an intrachromsomal misalignment of centromeric and telomeric LCR. Since NAHR could have occurred in any of the blocks (A, B or C), it can result in a variable range of sizes (from approximately 2.34 to 1.79 Mb (209).

II. Copy Number Variants

Apart from the paracentric inversion of 7q11.23, other susceptibility alleles predispose to rearrangements of the region. In 2008 it was found that approximately 4.44% of WBS-transmitting progenitors are carriers of a large deletion of LCR, compared to only 1% of control individuals. Duplications of LCR are also more prevalent in transmitting progenitors (approximately 2.22%) compared to controls (around 1.16%). All identified CNVs contained mostly pseudogenes and had no obvious phenotypic effect (219).

3.2 Williams-Beuren syndrome

WBS syndrome is a rare neurodevelopmental disorder that results from a heterozygous deletion of 26-28 genes at chromosome 7q11.23 and is estimated to affect approximately 1 in 7500 individuals (220). The first cases of WBS were described in the early 1950s (221), but it was not until 1961 when Williams et al, and a year later Beuren et al, suggested that the associated features represented a new syndrome (222,223).

3.2.1 Clinical features

WBS individuals show a set of shared clinical signs and symptoms, although the phenotype can vary across patients.

I. Facial features

Individuals with WBS have a distinctive facies characterized by a broad forehead, periorbital fullness, flat nasal bridge, short nose with a broad tip, long philtrum and a wide mouth (Figure 13). Young children are usually described as elfin, while adults tend to present coarser features, with a long neck and face (224).

Figure 13. Characteristic facial features of WBS patients. The pictures shows a female patient at different ages, from left to right: 2, 8 and 34 years respectively. From (225).

II. Cardiovascular Abnormalities

Cardiovascular disease is common among WBS patients and is the major cause of death among patients, with an associated mortality 25 to 100 times higher than in general population (226). Supravalvular aortic stenosis (SVAS) is present in approximately 70% of individuals and pulmonary arterial stenosis and mitral valve prolapse are also frequent (227). In addition, the frequency of hypertension among individuals with WBS is 50% and can begin as early as in childhood in some cases (224).

III. Endocrine abnormalities

Diabetes mellitus, subclinical hypothyroidism and hypercalcemia are the most frequent endocrine alterations seen in WBS. Impaired glucose metabolism is seen in a high percentage of individuals, with up to 75% showing abnormal results in oral glucose tolerant tests (228,229). Subclinical hypothyroidism is seen in 15 to 30%, often related to structural abnormalities of the thyroid gland (230,231). The frequency of hypercalcemia varies greatly (5 to 50%) across studies, possibly due to the differences in design and criteria.

Although it is more frequent in young children, adults with calcium alterations have also been reported (224).

Finally, children with WBS also show abnormal growth patterns, with a diminished adolescence growth spurt resulting in decreased adult height. In addition, females present an earlier menarche, on average two age earlier than general population (232,233).

IV. Other alterations

Gastrointestinal difficulties are also frequent among individuals with WBS. A frequent concern is constipation, which is present in around 50% of patients and can be a risk factor for diverticulitis or diverticulosis, which are also more prevalent in WBS population (233,234). Celiac disease is also more frequent and is found in approximately 10% of patients (235). Other gastrointestinal problems include feeding problems during infancy, gastroesophageal reflux and chronic abdominal pain (236,237).

Congenital renal genitourinary abnormalities in children with WBS are also frequent in children with WBS. The prevalence of renal anomalies in WBS is 17 times higher compared to general population and can range from minor anomalies to severe malformations, such as renal hypoplasia (238,239). These abnormalities result in a higher frequency of voiding dysfunction. The prevalence of abdominal wall or external genitalia abnormalities is also increased in WBS. Approximately 73% of males present anomalies, such as undescended testis, retractile testis, hypospadias or unilateral cryptorchidism. Females also have an increased frequency of anomalies, such as umbilical, unilateral inguinal or bilateral inguinal hernias (240).

WBS is also characterized by typical audiological anomalies, such as abnormal sensitivity to sounds and sensorineural or mixed hearing loss, mostly in the high frequency range that tends to worsen over time (241). Opthalmological anomalities are also frequent with most patients showing at least one ocular abnormality. The most frequent are strabismus and the presence of a typical stellate pattern of the iris (20-70%) and, less frequently, hyperopia, myopia, cataracts, astigmatism or tortuosity of renal vessels (234,242–245).

V. Cognitive and behavioral profile

Individuals with WBS present mild to moderate ID, with an average IQ of 55 to 62, that ranges from 40 to 102. However, a global measure of IQ is not representative of the distinct cognitive profile of people with WBS. Individuals with WBS show relative strengths in auditory memory and expressive language, but weaknesses in visuospatial construction and visuomotor skills (224,237,246,247).

Language acquisition is usually delayed in children with WS, both vocabulary and grammar onset. Interestingly, children with WBS show a delay in the use of referential pointing gestures, likely due to their deficits in visuospatial construction. Both expressive and receptive concrete vocabulary are conserved, whereas conceptual relational vocabulary is an area in which WBS individuals show difficulties. Although children and adults with WBS are usually very loquacious and show a good expressive language and correct grammar, individuals with WBS show difficulties with the pragmatic aspects of language. For example, they use an excessive number of stereotyped phrases or tend to show inappropriate over-familiarity (224,246,247).

Regarding their weak areas, WBS individuals present great difficulties in tasks involving visuospatial construction, such as pattern construction and drawing. When asked to draw an object, individuals with WBS focus on the parts rather than the global organization, resulting in poor cohesion and disorganization. However, drawing abilities seem to improve with age, consistent with the development process. Interestingly, performance on visuospatial tasks is directly correlated to performance on other areas, such as language, working memory and nonverbal reasoning (224,246,247).

As for behavior, patients with WBS have a friendly, empathic and hypersocial personality (246,247). For example, a study found that children with WBS show more interest in persons, even if they are strangers and in the presence of toys, compared to typically developing children (248). Moreover, individuals with WBS are more open towards strangers compared to children with other neurodevelopmental disorders or typically developing children. This could lead to dangerous situations, since individuals with WBS tend to approach strangers, disregard their countenance or possible social danger

clues. In contrast, individuals with WBS show difficulties in establish and maintain long term friendships (249).

Other concerns associated with WBS are psychiatric disorders, which have a high prevalence. Studies have reported that an important proportion of individuals with WBS (48-82%) met criteria for the diagnosis of one anxiety disorder (250–252). This rate is higher than among children with ID of other etiologies. Children with WBS also show a high rate of fears and of specific phobias (253). Also, one of the most frequent concerns of parents is the lack of motivation and pleasure in completing challenging tasks that can lead to low employment and independence levels in adults (254). ADHD has also a high prevalence among individuals with WBS (40-65%) (251,255).

3.2.2 Comorbidity with ASD

Due to their different associated behaviors, ASD and WBS have often been described as two diametrically opposite disorders (256), although some authors have suggested that this consideration is an oversimplification of both phenotypes (257). In this regard, several cases of WBS patients with co-existing ASD features have been described (258–269). Due to the rarity of WBS syndrome, most of these studies are anecdotal reports or are based on small sample sizes. Methods and criteria to evaluate features of ASD also vary across studies and difficult drawing conclusions about the frequency of ASD features among WBS patients (270).

However, a recent meta-analyses analyzed the presence of ASD features among several genetic syndromes, including WBS, selecting only high quality studies (270). The authors concluded that the prevalence of ASD features among WBS individuals is approximately 12%. Among the syndromes included, WBS was found to be among the group with lower prevalence of ASD symptoms. The study also found that syndromes with higher intellectual function, including WBS, had substantial lower incidence of ASD features. Still, the prevalence of ASD among WBS patients is approximately ten times higher than among general population, a striking finding especially considering the typical neuropsychological profile of individuals with WBS. Some authors have suggested that ASD should be reconsidered as part of the WBS phenotype (265).

Given that the WBS locus has been associated with mirror phenotypes in language and behavior, some authors have hypothesized that a trans-acting factor present in the remaining allele could act as a genetic modifying factor. In this regard, *GTF2I* and *GTF2IRD2*, the genes related to the cognitive phenotype, have been proposed as the main candidates. Another possible explanation is a cis-acting mechanism in which breakpoints of the deletion would alter flanking genes. However, the fact that nearly all of the patients described present the common deletion excludes this possibility (266). Finally, a polygenic model in which interaction of genes present in the WBS locus with the genetic background could modify the expression of the phenotype has also been proposed (265,266). Some authors have observed that some WBS-ASD patients, present hyperserotonemia, which is also a frequent finding among individuals with ASD of another etiology (261,266). For that reason, the serotonin transporter gene *SLC6A4* has been proposed as a possible modifier factor (266). However, no unbiased genome-wide studies have been carried out to study the contribution of other genes.

3.3 7q11.23 duplication syndrome

The first case of a 7q11.23 duplication was described in 2005 (213). Since then, 46 children and 15 adults with the syndrome have been reported (215). Some of the patients have been identified by CMA in large scale studies of various neurodevelopmental or psychiatric conditions (214,271–274), revealing that the duplication is a risk factor for ASD and schizophrenia (58,275). Overall, it seems that the duplication phenotype is much more variable than its reciprocal deletion.

3.3.1 Clinical features

I. Facial features

Although not so recognizable as the typical facial features of individuals with WBS, carriers of the 7q11.23 duplication also share some common craniofacial features. Common facial features include macrocephaly (approximately in 50% of individuals), brachycephaly, broad forehead, straight eyebrows, broad nasal tip, low insertion of the columnella, short philtrum, thin upper lip and facial asymmetry (Figure 14) (214,215,276).

Figure 14. Characteristic facies of individuals with the 7q11.23 duplication. The pictures show five young children with the classic 7q11.23 duplication.

II. Cardiovascular Abnormalities

Cardiovascular abnormalities are a frequent finding among patients with 7q11.23 duplication. The most frequent abnormality is dilation of the aorta, which has been described in 46% of patients of all ages in large series. Congenital cardiovascular malformations have also been described, including septal defects, subvalvar aortic stenosis and patent ductus arteriosus (215,277).

III. Abnormalities affecting other organs and systems

Gastrointestinal anomalies are also frequent, with chronic constipation affecting more than half of children (around 66%), than, in some cases, need medical intervention (215).

Similar to WBS, genitourinary anomalies have also been observed in patients with the reciprocal duplication. They affect around 20% of patients and range from unilateral renal agenesis to hydronephrosis (215,278).

Various musculoskeletal alterations have also been noted in approximately 28% of patients with the 7q11.23 duplication, such as hyperextensible joints, pectus excavataum, lordosis or pes planus (215).

IV. Cognitive and behavioral profile

The cognitive profile of the 7q11.23 duplication is usually contrasted to that of WBS. Individuals with the 7q11.23 present difficulties with verbal skills and in social interaction (279,280). This was evident from initial case reports, but it was not until recently that individuals bearing the 7q11.23 duplication were systematically evaluated in order to define their cognitive and behavioral profile (279).

Regarding cognition, individuals with the duplication were found to have a Differential Ability Scale (DAS) score, which is similar to IQ, of 85, ranging from severe disability to high average hability. In toddlers and adults, a pattern of relative strength in non verbal reasoning and poorer performance in verbal abilities could be observed. Adults also showed a relative strength in visuospatial construction (279).

As for language, around 82% of children with the duplication meet criteria for speech sound disorder, according to DSM-5. Speech delay is also a frequent finding among affected children. The most common subtypes of speech sound disorder are dysarthria and childhood apraxia of speech. The percentage of children with a speech sound disorder seems to be lower (40%) among adolescents, although most of them had received or still received speech teraphy (279,280).

Regarding behaviour, children with the duplication are often described as shy and withdrawn when interacting with unknwon people. An important percentage of individuals with the duplication, were diagnosed with various disorders. For instance, 50% met criteria for social phobia, 29% for selective mutism, 13% for separation anxiety disorder and 6.5% for generalized anxiety disorder. ADHD is also frequent, with 35% of children diagnosed. Although anxiety disordres are present in both WBS and the reciprocal duplication, the duplication seems to increase the risk for social phobia and selective mutism, which has not been reported in WBS (279,280).

3.3.2 Comorbility with ASD

Although several case reports have identified children with duplication meeting criteria for ASD diagnosis, no formal assessment has been carried out in a large sample to asses the frequency of ASD in individuals with the duplication. However, screening using the Social Communication Questionnaire (SCQ) showed that around 30% of children had a positive screening (279). A formal assessment with clinically validated tools, like ADI-R and ADOS is needed to confirm or discard the diagnosis and establish the frequency of ASD among individuals with the duplication. Some authors have argued that, although children with the duplication present social anxiety and speech problems, they also show characteristics that are not compatible with the diagnosis of ASD. For example, they maintain adequate eye contact with

their peers and enjoy interacting with their partners and imaginative play (279). However, given the frequency of the duplication among children with ASD, the 7q11.23 could represent a risk factor for the disorder ($p<0.003$) (58,279,280).

As seen in the previous section, individuals with reciprocal deletions and duplications of the 7q11.23 present mirror phenotypes in some clinical aspects (Table 5).

	7q11.23 deletion	7q11.23 duplication
Facial characteristics	Broad forehead Low nasal root Long philtrum Full lips	Broad forehead High, broad nose Short philtrum Thin lips
Growth/endocrine problems	Growth retardation Hypercalcemia	Normal growth Normal calcemia
Cardiovascular abnormalities	SVAS	Dilatation of the aorta
Connective tissue anomalies	Joint laxity	Joint laxity
Cognitive profile	Developmental delay ID Relative strength in expressive language Deficit of visuospatial skills	Developmental delay ID Speech and language delay Visuospatial skills conserved
Behavioral problems	Hypersocial personality Increased prevalence of ASD ADHD	Social deficits Increased prevalence of ASD ADHD

Table 5. Comparison of clinical characteristics between patients with deletions or duplications of the 7q11.23 region. Adapted from (219).

The increased prevalence of ASD features in both types of rearrangements (Table 5) suggests that the 7q11.23 region plays an important role in behaviour, language and cognition. Therefore, the detailed study of this region can provide insights into genes and mechanisms altered in neurodevelopmental disorders. However, the study of this region is difficulted by its complex genetic architecture, since one of the main candidate genes for the cognitive profile, *GTF2I*, lies in a duplicated region.

3.4 The role of the transcription factor II-I family in cognition

Several approaches have been used to study the specific contribution of genes in the WBS locus to the features of the WBS and the reciprocal duplication. These strategies are mainly the comparison of clinical and molecular characteristics of patients with typical and atypical duplications, mouse models and functional gene studies. Evidences from these studies point towards a major role of the transcription factor II-I family (TFI-II) family of factors in cognition and behavior (281).

The TFII-I family contains three genes (*GTF2I*, *GTF2IRD2* and *GTF2IRD1*) with multiple helix-loop-helix-like domains (known as I-repeats). *GTF2I* has a ubiquitous expression, with wide expression in the developing head. It is thought to act as a transcription factor that can bind to various enhancer and core promoter elements (282). Interestingly, *GTF2I* was found to be paternally imprinted, pointing towards an epigenetic mechanism regulating its expression (283).

Genomic alignments suggest that *GTF2IRD2* may have evolved from *GTF2I*, and that it may be a truncated version (284,285). Whereas *GTF2I* is deleted in both the 1.55Mb common deletion and the large 1.8Mb deletion, *GTF2IRD2* is deleted in only the larger 1.8Mb deletion (209). Although *GTF2IRD2* has three copies, located in the Bc, Bm and Bt blocks, only the copies present telomeric and medial blocks are fully transcribed, since the centromeric copies lack the first two exons (Figure 10) (286). *GTF2IRD2* interacts with *GTF2I* and may act as a regulator by sequestrating *GTF2I* to inactive nuclear regions (287).

The last member of the TFII-I family is *GTF2IRD1*, which is located in the single copy region of the WBS locus. It is a regulatory factor that binds several genes with a highly conserved DNA element (288). *GTF2IRD1* is also highly expressed during development (289).

GTF2I is the main candidate gene related to the neurocognitive profile seen in WBS patients. Evidences come from patients with atypical deletions, as well as mouse models. On the one hand, several case reports of patients with smaller atypical deletions that do not include *GTF2I* showed a milder cognitive phenotype (with normal IQ in some cases) and no hypersocial personality. They also implicated *GTF2IRD1* in the visuospatial construction defects seen in most patients with atypical deletions (205,290–292). On the

other hand, a patient with an atypical deletion including the three members of the TFII-I family was also described. The patient, who met criteria for ASD but also presented the specific WBS cognitive and behavior profile, further suggested the role of this gene in the cognitive and behavior deficits (263). Finally, evidences from atypical rearrangements also come from patients with atypical duplications. A study of CNVs among the Simons Simplex Collection (SSC) screening for genetic factors for ASD, revealed a *de novo* duplication in a patient, encompassing the interval from *GTF2IRD1* to *WBSCR16*, which also includes *GTF2I* (293).

Regarding mouse models, in 2010 a mouse model that lacked the first the first 140 aminoacids of *Gtf2i* was described. This model recapitulated, in heterozygosis, some of the features present in WBS, such as craniofacial alterations, sound intolerance or increased anxiety ((294). A second mouse model described in 2011, which was also deficient in *Gtf2i,* showed an hypersocial behaviour, reinforcing its role in the hypersocial personality (206).

Regarding *GTF2IRD2*, a study comparing WBS patients with the typical 1.55Mb deletion with patients with the larger 1.8Mb (which includes *GTF2IRD2*) found that the second group presented more cognitive impairment and a higher rate of externalizing behaviours. This suggests that GTF2IRD2 may influence reasoning and executive function (207).

4. Genetic counseling

Information regarding genetics and inheritance in the context of genetic diseases can be difficult to understand and to assimilate. However, it is key to adapt to a risk or a condition and to take informed choices. For that reason, there is a need for genetic counselors, professionals specialized in medical genetics and counseling that can provide genetic counseling.

Several definitions of genetic counseling have been proposed over the years, but the most recent and widely accepted is the 2006 definition proposed by the National Society of Genetic Counselors (NSGC) (295). According to the NSGC, genetic counseling is "the process of helping people understand and adapt to the medical, psychological, and familial implications of the genetic contributions to disease. This process integrates: 1) Interpretation of family and medical histories to assess the chance of disease occurrence or recurrence; 2) Education about inheritance, testing, management, prevention, resources and research; and 3) Counseling to promote informed choices and adaptation to the risk or condition" (295).

Accordingly, genetic counseling aims to assist affected and/or at risk individuals to understand the nature of a genetic disorder, its transmission and available options for management and family planning, as well as to cope with derived psychological stress and minimize it. Families and patients use the information provided during a genetic counseling session to make decisions related to their genetic condition and according to their risk, family goals and ethical or religious convictions. In the context of common disorders, where environmental factors play a substantial role in the development of the disorder, genetic counseling can help individuals to adopt healthy habits and to reduce the risk of developing a disease. Since genetic diseases are heterogeneous, genetic counseling is necessary and present in a variety of settings: assisted reproduction, cardiovascular disease, familial cancer, personalized medicine, psychiatric disorders, reproductive and prenatal counseling, etc (296).

The goals of genetic counseling are both educational and psychosocial. According to this, two main models for genetic counseling have been proposed: a teaching model and a counseling model. On the one hand, the teaching model is based on the assumption that patients come for

information and, therefore, education and information giving are considered the central parts of the process (297–299). On the other hand, the counseling model emphasizes a psychotherapeutic approach, where the emotional well being of patients is a key part of the process (298,299). Some authors have suggested that genetic counseling should be based on a combination of both models and apply both skills depending on the individual's needs (296,298,300,301).

One of the principles of genetic counseling is non-directiveness, which aims to increase the autonomy of the individual, as well as his or her sense of personal control (296). To that end, the genetic counselor should not only provide comprehensive and understandable information, but also use his or her counseling skills to increase the individual's self-directness (302). For example, genetic counselors should increase the sense of autonomy of the patient, by validating their feelings and acknowledge their strengths and competencies (302). It is also important to guide patients to their own decisions by helping them to think through their situations. One of the context in which non-directiveness acquires more relevance is in the prenatal setting, where individuals are often faced with the decision of terminating a pregnancy. In this context, genetic counselors could help a patient to reach a decision by asking how she would feel in each situation, providing a framework by which she could think through all the aspects of the situation and reach her own conclusion (302). In addition to the prenatal setting, non-directiveness is also relevant in other contexts, such as pre-symptomatic testing.

Besides counseling abilities, communication skills are also essential to provide good genetic counseling in a non-directive manner. In order to take informed choices, individuals need to attain a good understanding of their risk. Some studies have shown that the manner in which risk is communicated can influence patient's perceptions (303,304). It is also important to know factors that can influence its understanding and its subjective perception, such as education, prior experiences or preconceptions about the severity and the risk (305). Some recommendations about communicating risks in an effective manner have been published (306,307). They include informing about the baseline population risk to allow for comparison, framing the risk both in a positive and

negative manner and using figures and graphs. It is also very important to check the counselee's understanding by asking questions and allowing time to assimilate the information (306,307).

Genetic counseling is essential in complex disorders, such as ASD. Although ASD has a strong genetic component, in most families a genetic cause cannot be identified. Therefore, most families with one or more children with ASD are faced with great degree of uncertainty, especially when considering family planning and their reproductive options. Genetic counseling can help families understand the genetic component of ASD, their recurrence risk and to cope with the disorder.

4.1 Genetic counseling in ASD

Genetic counseling in ASD is complicated by the great genetic heterogeneity and the low diagnostic yield of available genetic tests. Current tests applied into the clinical setting (karyotype, X-fragile testing, single gene testing and CMA) have a combined diagnostic yield of approximately 20-30% (46). Recent technologies, such as exome sequencing, are expected to increase the diagnostic yield by at least 10% (159), but are still not applied to the clinics and studies evaluating their acceptance among families are lacking.

In genetic counseling, it is key to establish an accurate diagnosis in the index case or, if not possible, determine the inheritance pattern of disease in the family in order to offer appropriate RR. In addition to that, genetic diagnosis is useful to offer information about the prognosis of the disorder and monitor and prevent associated medical conditions. When a specific etiology is identified, genetic counseling and recurrence risk can be provided on the basis of that specific diagnosis. All possible patterns of Mendelian inheritance, except Y-linked inheritance, have been described in ASD: autosomal dominant or recessive and X-linked dominant and recessive conditions. In that case, recurrence risk is usually straightforward.

However, in most cases of ASD, the etiology remains unknown and families must base their reproductive decisions on recurrence risk estimates based on empirical data. Several studies evaluating the recurrence risk have been performed, with different recurrence rates among full siblings (see Figure 15 for a general overview) (44,45,308–314).

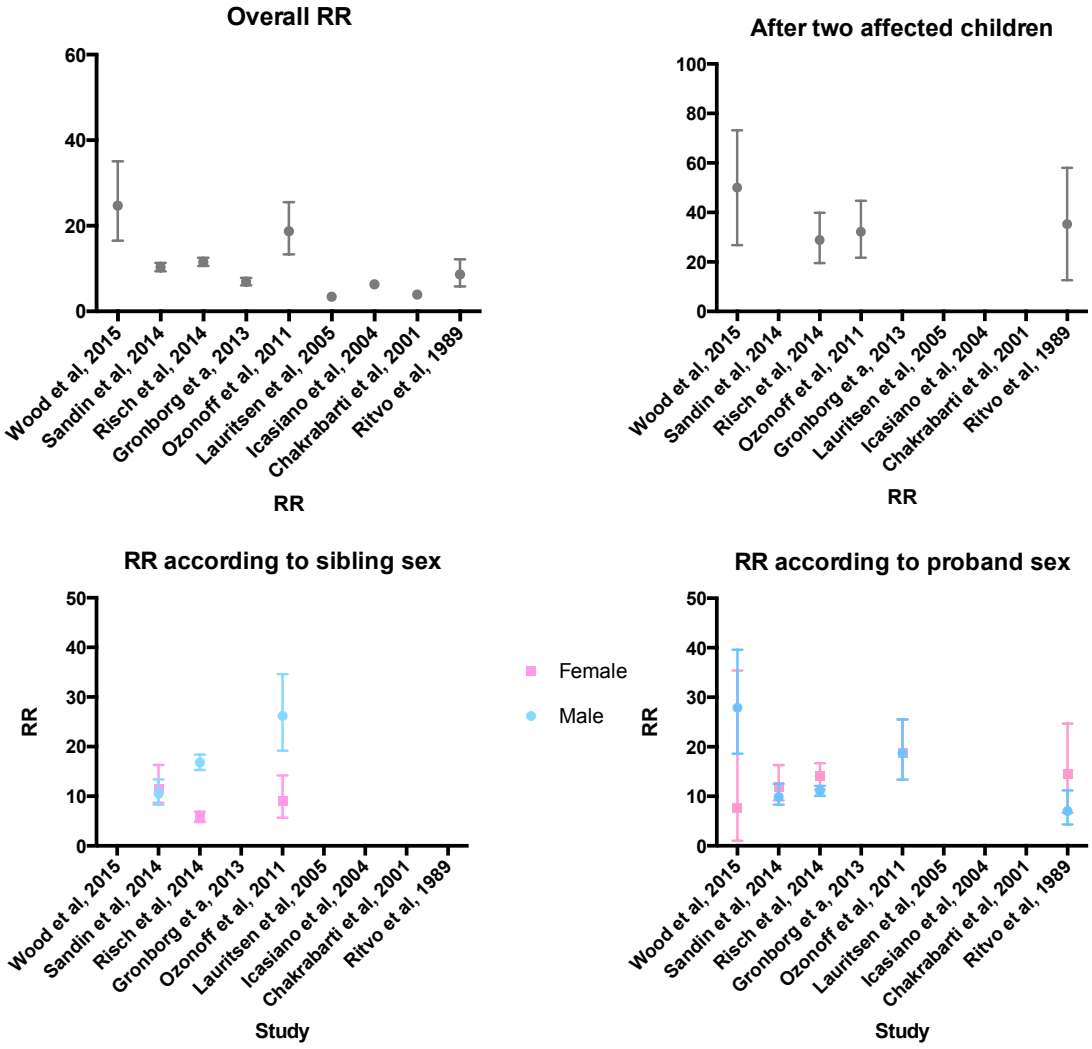


Figure 15. Recurrence risk (RR) estimates for ASD in several epidemiological studies. Estimated recurrence risk in different situations is depicted with 95% confidence interval (CI) bars. Studies not examining specific recurrence risk are left for comparison, but no data is depicted. The upper part of the figure shows recurrence risk estimates in all families (right) and in families with more than one affected child (left) for both sexes combined. The lower left part shows specific risk estimates separated by sex of the sibling (blue for male siblings and pink for female siblings). In the lower right part of the figure, the recurrence risk for siblings of male probands is shown in blue and for female probands in pink.

Earlier studies placed the recurrence risk of ASD between 3 and 7% (308–311), and in the clinical practice a conventional and general risk figure of 5-10% is usually given (315). However, more recent studies in larger cohorts point towards a higher RR, estimated to be between 10 and 24% (44,45,312–314). This risk may even be higher for the BAP, with studies pointing towards

a 17.5% frequency among non-affected siblings (314). Therefore, quoting a recurrence risk of 10-20% may be a more accurate estimate. The number of affected children should also be taken into account when communicating RR. Some studies have also estimated recurrence risk in families with more than one child affected, showing that the risk is higher in those families, approximately 30-50% (44,313,314).

Results between studies also differ in sibling sex liability to ASD. Two studies suggest that recurrence risk may be higher for male siblings than for female siblings (44,313), although these results are not consistent in all studies (45) (Figure 14). Finally, sex of the affected child could also have an effect on recurrence risk for siblings of both sexes, with higher recurrence risk in siblings of female probands (Figure 14). This is consistent with a multifactorial model of inheritance, in which risk is increased in the relatives of a proband of the less commonly affected sex. However, there is mixed evidence for this model, since studies are not consistent in this point (Figure 14).

One of the possible and suggested explanations for the lower recurrence risks of earlier studies is their inability to account for reproductive stoppage. Reproductive stoppage is a frequent phenomenon among families with a child affected with ASD, where parents decide to have none or fewer children after the first affected child. Various studies have shown that reproductive stoppage is extended among families with children with ASD (308,314,316,317). Corroborating this finding, one of the studies showed that reproductive behavior of affected families does not change for the first years after the first affected child is born, a time when the diagnosis has not been established. However, after that, the birth rate of families with an affected child falls compared to control families (317).

Genetic counselors should consider this body of knowledge when communicating recurrence risk to parents. They should also take into account that the degree of severity is unpredictable, whatever the diagnosis of the first child. Importantly, the most reliable and useful tool to predict recurrence risk is still family history. Therefore, it is essential to properly collect clinical information and correctly draw the family tree, which can help to identify a pattern of inheritance and other family members at risk. In the case of ASD,

clinicians should pay special attention and specifically ask for other neurodevelopmental and psychiatric disorders in the family, as well as for BAP features in the extended relatives. The genetic counselor should consider these additional risk factors present in the family history to adjust the recurrence risk in each individual case (318). For instance, in the presence of a clear and identifiable pattern of X-linked inheritance, specific and Mendelian risk figures should be given instead of empirical estimates, even if the underlying genetic cause can not be identified.

Besides numeric recurrence risk estimates, other factors can influence perceived recurrence risk and acceptance of genetic testing, such as perceived causes of etiology and motivations for genetic testing. All these factors should be explored during genetic counseling sessions, since they can affect decision-making processes. Several studies have demonstrated that perceived recurrence risk has an effect on family planning and that genetic services are underutilized among families with children with ASD (319–326). However, all studies published up to date have been carried out in USA, where the health system varies greatly from the Spanish's public health system. Therefore, results could differ from those in European countries, due to factors related to insurance polices. In addition, cultural factors may also shape perceived causes of ASD and RR.

Finally, genetic counseling should also take into account psychosocial factors, such as the impact that raising a child within the spectrum has on family dynamics and possible derived feelings of guild among parents (319,327–330). Although few data is available about recurrence risk in extended family members (45), such as cousins or siblings of affected individuals, genetic counseling should also be offered to extended family members. When counseling for ASD, it is important to acknowledge that some families and individuals do not see ASD in a negative light, but rather as the result of a different developmental process. Therefore, some of the terms usually used in the clinical practice can carry negative connotations and provoke negative reactions. To avoid this, terms such as *risk* or *normal* development can be replaced for more neutral words, such as *probability* or *neurotypical* development. Feelings of self-identity and community belonging have been widely reported in individuals with Asperger syndrome, who were especially concerned about the latest changes in classification in DSM-5

(331–333). Also, contact to patients and family associations should be offered to parents, since they can offer them shared experiences and ongoing support.

OBJECTIVES

Objectives

The aim of this thesis is to contribute to the knowledge of the genetic architecture of ASD, as well as to provide insights into deregulated mechanisms leading to the disorder. To conduct this work, we relied on the indispensable collaboration of approximately 300 families affected with the disorder, who generously gave samples and access to their detailed family history and clinical characteristics.

In order to achieve this global aim, we set the following specific objectives:

1. To identify pathogenic variation by the join study of exome and transcriptome data in a group of males with idiopathic ASD, in order to pinpoint loci involved in the etiology of the disorder.

2. To study the contribution and the extent of previously identified loci in a larger group of individuals, in order to reinforce their role in the disorder.

3. To explore the variation and the role in ASD of two complex regions of the genome specifically duplicated in the human lineage, the 7q11.23 and the pericentromeric region of chromosome 9, which have been neglected by previous studies.

4. To search for second-hit genetic factors that may explain the co-occurrence of ASD in a group of individuals with WBS, in order to identify ASD susceptibility loci of milder effects.

5. To study perceptions, knowledge and the effect of genetic counseling in a group of parents with children affected with ASD in order to provide appropriate genetic counseling to families, as well as to contribute to the translation of research findings.

CHAPTER 1

Marta Codina-Solà, Benjamín Rodríguez-Santiago, Aïda Homs, Javier Santoyo, Maria Rigau, Gemma Aznar-Laín, Miguel del Campo, Blanca Gener, Elisabeth Gabau, María Pilar Botella, Armand Gutiérrez, Guillermon Antiñolo, Luis A. Pérez-Jurado* and Ivon Cuscó*
Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders.
Mol Autism. 2015 Apr 15;6:21.

CHAPTER 2

Targeted sequencing of Autism Spectrum Disorder candidate genes identifies pathogenic mutations and an excess of rare variants in the neurexin and neuroligin superfamily of genes

Marta Codina-Solà, Gaudí Valls, Gemma Aznar, Teresa Vendrell, Anna Cueto, Miguel del Campo, Elisabeth Gabau, Luis A. Pérez-Jurado, Ivon Cuscó

In preparation

Abstract

Introduction: Despite the strong genetic component of Autism Spectrum Disorders (ASD), the etiology of most than 70% of cases remains unknown. Genetic factors involved in ASD affect hundreds of genes, making it difficult to prove the causality of new candidate loci. Despite this high degree of genetic heterogeneity, altered genes cluster in functionally related pathways, such as synaptic cell adhesion molecules. To establish the role of candidate genes identified by a previous exome sequencing study, as well as to study the contribution of rare mutations in the neurexin and neuroligin gene superfamilies, we performed a case-control study of 38 candidate genes.

Materials and methods: We selected a group of 38 genes on the basis of previous evidence from our group or their association to synaptic cell adhesion molecules and sequenced their coding regions in 279 ASD patients and 105 population-matched controls. We applied a comprehensive annotation and filtering pipeline to identify variants of high penetrance and perform gene and pathway-based analysis to identify sets of rare variants associated to ASD.

Results: We identified *de novo* pathogenic mutations in various genes of high penetrance (*SCN2A*, *CDKL5*, *SHANK3* and *CNTNAP2*), as well as inherited pathogenic variants in two genes (*CNTNAP2* and *NRXN2*). Although gene-based association studies yielded no significant results, cases showed a statistically significant higher burden of rare variants in loss of function intolerant genes related to synaptic functions.

Conclusions: Our results demonstrate the utility of pathway-based approaches in the study of complex disorders with high genetic heterogeneity such as ASD. Overall, this work illustrates the complex genetic architecture of ASD, with both *de novo* mutations and inherited variants of milder effects contributing to risk.

Introduction

Autism Spectrum Disorders (ASD) are a group of neurodevelopmental disorders of early onset, defined by communication deficits, impaired social interaction and a restrictive and repetitive pattern of behavior. ASD comprise a heterogeneous group of complex disorders and their manifestations can vary greatly among individuals. They have a strong genetic component, as shown by high heritability estimates and increased recurrence risk among siblings of affected individuals (1–10). However, a genetic cause can be identified in only 20 to 25% of cases. Genetic factors involved in ASD comprise all type of genetic variation and affect hundreds of genes, reflecting its high degree of clinical heterogeneity (11).

Despite their high degree of genetic heterogeneity, several evidences point towards the role of functionally related genes, such as synaptic cell adhesion molecules and synaptic scaffold proteins, in the etiology of ASD. This signaling pathway includes several gene families, such as neurexins (NRXNs), neuroligins (NLGNs), contactin-associated proteins, contactins and Shank family proteins. NRXNs and NLGNs are membrane bound molecules that mediate synapsis, connecting pre and postsynaptic neurons, respectively. The human genome contains three NRXN (*NRXN1*, *NRXN2* and *NRXN3*) and four NLGNs (*NLGN1*, *NLGN2*, *NLGN3* and *NLGN4*), all of which have been associated to ASD (12–20). In addition, some members of the contactin-associated proteins (*CNTNAP1*, *CNTNAP2*, *CNTNAP3*, *CNTNAP4* and *CNTNAP5*), which belong to the neurexin superfamily of genes, have also been extensively related to ASD (21–26). Contactin-associated proteins interact with contactins, which mediate synaptic plasticity and glia interactions (27) and are also altered in ASD (28,29). Finally, mutations in synaptic scaffold proteins from the Shank family, such as *SHANK3*, have also been described in patients with ASD (30–33).

Besides synapse-related genes, recent high-throughput studies in ASD have uncovered novel candidate genes, associated to transcriptional regulation and chromatin-remodelling pathways. Most studies have followed a trio-based design and focused on *de novo* variation to identify recurrently mutated genes (34–41). However, recent approaches have also studied the burden of inherited variation (42–44). Although *de novo* mutated genes can give an insight into new deregulated pathways*, de novo* mutations occur

randomly across all the genome, with an estimated average of approximately 74 Single Nucleotide Variants (SVNs) per genome and per generation (45). Therefore, identifying a single *de novo* event in a gene does not necessarily implicate pathogenicity and recurrent events or functional evidences are needed to confirm its role. Thus, due to the great genetic heterogeneity of ASD, some genes may remain elusive and require replication studies to identify additional cases.

In a previous exome sequencing study, we identified several candidate genes that could influence the risk of ASD by following a monogenic or oligogenic model of inheritance (46). Several genes presented variants of high penetrance, either *de novo* or X-linked (*CDKL5*, *CUL3*, *MED13L*, *KCNV1*, *SCN2A* and *MAOA*). In addition, we identified a group of genes with inherited variants of milder penetrance that could act as susceptibility factors in a multiple hit model. This group consisted of recurrently mutated genes (*ANK3*, *ASMT*, *NLRP8*, *RBM19* and *SVEP1*) and genes harboring loss of function (LoF) variants in our cohort but none or very few in data from Exome Variant Server (*ALG9*, *DDHD2*, *IGSF10* and *RTN3*). To establish the role of candidate genes previously identified by exome sequencing, as well as to study the contribution of rare mutations in the neurexin and neuroligin gene superfamilies, we performed targeted sequencing of 38 genes in 279 ASD patients and 105 controls. We study the contribution of highly penetrant variants by performing segregation studies of deleterious variation, as well as the burden of rare variants by gene-based and pathway-based case-control association studies. In addition, we explore the contribution of other sources of variation, such as somatic SNVs and small exonic Copy Number Variants (CNVs).

Materials and methods

Sample selection

We studied 279 unrelated patients with a diagnosis of idiopathic ASD selected from a European cohort of 519 affected individuals. All had a confirmed diagnosis of one of the categories of ASD listed in the Diagnosis and Statistical Manual of Mental Diseases IV (DSM-IV), categorized

according ADI-R (Autism Diagnostic Interview-Revised. The study was approved by the Clinical Research Ethics Committee of the involved centers (CEIC-Parc Salut Mar), and informed consent for participation was obtained from the parents or legal caregivers. Parental and familial samples were obtained from available relatives who gave informed consent.

Control samples (n=105) were randomly selected from a cohort of Spanish anonymous blood donors (*n* = 600). Genomic DNA was extracted from peripheral blood by the salting-out method using the Puragene DNA purification Kit (Gentra Systems).

Targeted sequencing probe-set design

A total of 38 genes were selected, based on evidence from a previous exome sequencing study (n=15) or on their association to synapse-signaling pathways (n=23) (Supplementary Table 1) (46). The first group included genes with previously detected *de novo* and X-linked mutations (n=6), as well as possible risk loci: recurrently mutated genes (n=5) and genes harboring LoF variants in patients (n=4), but none or very few in publicly available data from Exome Variant Server (EVS) (46,47). LoF variants were defined on the basis of their functional effect and included stop, frameshift and splicing mutations. The second group of genes consisted of 23 genes from various gene families: neurexins, neuroligins, contactins, contactin-associated proteins and members of the Shank family.

Library preparation, capture and sequencing

DNA was fragmented using a Covaris ultrasonicator (Covaris, Woburn, USA) to 200-400 bp fragments. Libraries were prepared following Illumina's TruSeq standard protocol (Illumina, Inc., San Diego, CA, USA) and double-indexed samples were pooled together (96 samples per pool). Target genes were captured using a custom design panel by NimbleGen SeqCap EZ custom Library (Hofman-La Roche, Basel, Switzerland). Finally, enriched fragments were paired-end sequenced at a read length of 150 bp on an Illumina MiSeq (Illumina, Inc., San Diego, CA, USA).

Sequence analysis pipeline, variant annotation, and prioritization

Quality control was assessed using FastQC and adaptor sequences were trimmed using Trim Galore. Sequences were aligned to the latest version of the human genome (hg19) using the Burrows-Wheeler aligner (BWA) with default parameters, PCR duplicates were marked and removed using PicardTools, and quality scores of alignments were recalibrated using the GATK toolkit. Single nucleotide variants (SNV) and indel calls were only considered if positions had a depth of coverage of at least 10×, and heterozygous positions were only called when a minimum of 20% of the reads showed the variant (AB between 0.2 and 0.8). Annotation of variants was performed using ANNOVAR (http://www.openbioinformatics.org/annovar/), taking into account the variant frequency in control databases: dbSNP137 (http://www.ncbi.nlm.nih.gov/SNP/), Exome Variant Server (EVS) (http://evs.gs.washington.edu/EVS/), Exome Aggregation Consortium (ExAC), Database of genomic known variants (Kaviar) and an in-house database of 90 Spanish controls (48). The nature of the changes was assessed by SIFT, PolyPhen and Condel protein effect prediction algorithms (http://bg.upf.edu/fannsdb/). To distinguish putative disease-causing variants, we selected unique exonic variants and those affecting canonical splicing sites. Synonymous variants and variants previously described in the general population (ExAC and Kaviar) with a minor allele frequency (MAF) >0.002 were filtered. Variants selected for validation were manually inspected to exclude false positives using Integrative Genomics Viewer (IGV).

CNV detection

XHMM was used to call CNVs, following the standard steps as described in the online tutorial. We applied the same filters previously described (cita): XHMM quality score (SQ) ≥65, exons spanned ≥3, and estimated CNV length ≥1kB. We focused our analysis on rare CNVs, so we excluded CNVs overlapping with polymorphic variants reported in Database of Genomic Variants (DGV) (http://dgv.tcag.ca/dgv/app/home).

Detection of mosaic SNVs

In order to detect mosaic SNVs at lower AB ratios (<0.2), we performed variant calling using MuTect2. Variants were filtered following Mutect's default hard filtering. To discard non-somatic events, we filtered recurrent and

reported variants in the previously mentioned control databases. Finally, we selected exonic variants and manually revised all variants to exclude false positive using Integrative Genomics Viewer (IGV).

Validation

To perform validation and segregation studies in selected variants, we used Sanger sequencing by capillary electrophoresis (ABI PRISM 7900HT, Applied Biosystems, Foster City, CA, USA). PCR reactions were performed under standard conditons. For CNV validation, we used multiple ligation probe amplification (MLPA) with custom probes in the target region. MLPA reactions were carried out under standard conditions. The relative peak height method was used to determine the copy number status.

Paternity testing

To corroborate paternity on patients with *de novo* mutations, we performed microsatellite genotyping parental and patient samples. We selected highly heterozygous microsatellites markers randomly distributed in different autosomal chromosomes. PCR products were amplified under standard conditions, and fragments were separated and analyzed by high-resolution electrophoresis using GeneMapper software (ABI 3100, Applied Biosystems, Foster City, CA, USA.

Statistical analysis

Allele frequencies in ASD patients and control were compared using Pearson's $\chi 2$ and Fisher's exact tests. Statistical analysis was performed using R-platform software. A P-value < 0.05 denoted the presence of statistically significant differences.

Results

Overall, we identified 2214 different variants, with a mean of 198 variants per individual. To prioritize disease-causing variants, we filtered variants based on frequency and predicted pathogenicity. Since ASD is a highly genetic heterogeneous disorder, we excluded common (MAF > 0.002) and non-unique variants. To distinguish rare innocuous variation from those with

potential pathogenic effect, we focused on variants with possible functional consequences that affected either exonic or canonical splicing sites (Figure 1). After filtering, we remained with a total of 322 different variants, with an average of 1 variant per individual.

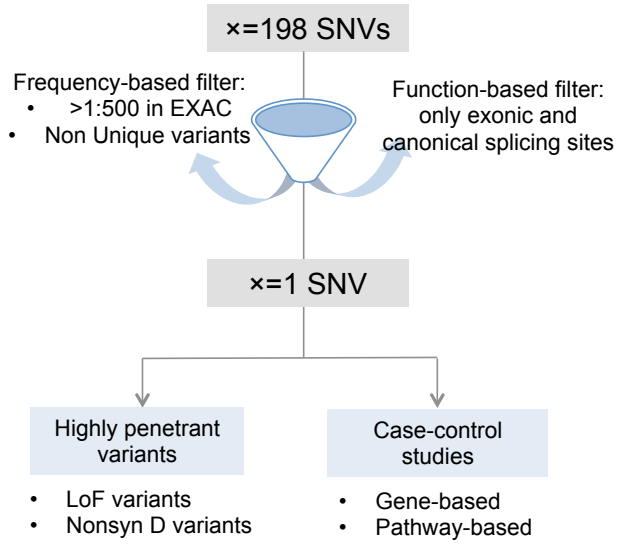


Figure 1. Prioritization pipeline used to select variants for segregation and case-control studies studying the burden of rare variants.

Highly penetrant variants

In order to identify highly deleterious SNVs, we selected LoF and non-synonymous variants predicted as damaging according to SIFT and Polyphen prediction algorithms. We identified five LoF variants in patients and one in controls, as well as 20 nonsynonymous damaging variants in patients and six in controls (Table 1).

	Candidate genes (n=15)			**NX, NL and contactin gene families (n=23)**		
	LoF	**Nonsyn D**	**Total**	**LoF**	**Nonsyn D**	**Total**
Cases (n=279)	2	7	9/279	3	13	16/279
Controls (n=105)		3	3/105	1	3	4/105

Table 1. Prevalence of novel deleterious variants in control and cases.

Validation and segregation studies were performed in all available trio samples (n=6) (Supplementary Table 2). All variants (n=9) were validated and three were found to be *de novo*. Paternity was confirmed in all cases. The three *de novo* variants were two nonsynonymous mutations affecting highly conserved residues in *CNTNAP2* (p.S820R) and *CDKL5* (p.Y171C) and a stopgain mutation in gene *SHANK3* (p.Q1275X).

Two LoF variants were detected in individuals with no available parental samples. The first was a stopgain variant in *SCN2A* predicted to result in a prematurely truncated protein lacking the third and four transmembrane domains. Although no parental DNA was available, it is probably a *de novo* variant, since no LoF inherited variants have been described in . The second LoF variant affects a canonical splice site of *CNTN2* (c.3014-2A>G) and is found in an anonymous healthy control. However, the probability of LoF intolerance according to ExAC in this gene is very low (pLI=0.13).

Among inherited variants, we identified a paternal inherited nonsense mutation in *CNTNAP2* (p.W184X*)*, and two maternally inherited LoF mutations in *RBM19* (c.2196_2203del) and *CNTN6* (p.Q740X). Interestingly, the nonsense mutation in *CNTN6* was found in the same patient carrying the *de novo* stopgain mutation in *SHANK3* (p.Q1275X). Finally, we identified an inherited novel missense mutation in *NRXN2* affecting a highly conserved residue (p.R728C), which segregated with the phenotype in a multiplex family. The variant was found in two affected siblings, but not in a third unaffected sibling, who presented specific language impairment (SLI). It was inherited from the father, who was diagnosed with Obssessive-Compulsive Disorder (OCD) (Figure 2).

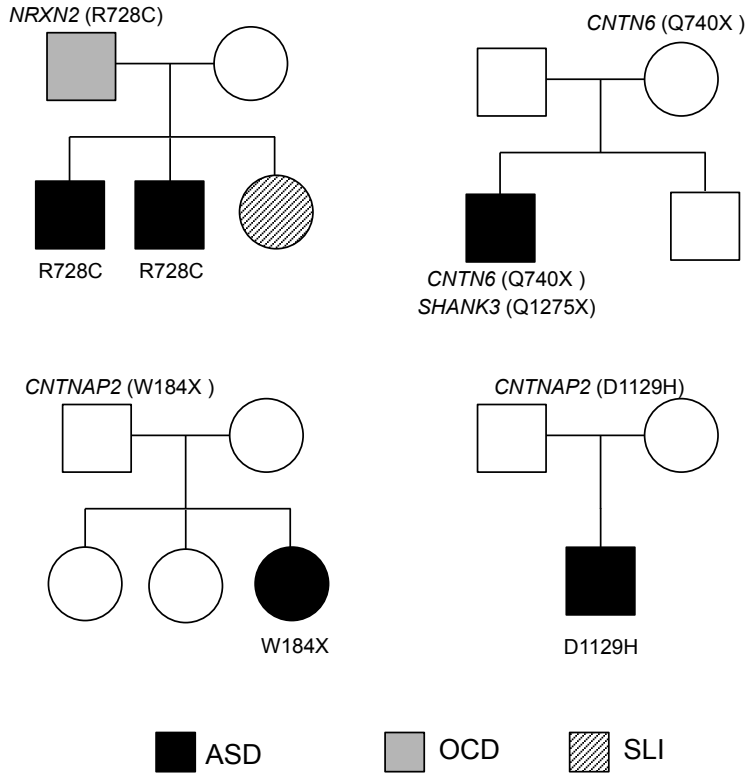


Figure 2. Inherited possible pathogenic rare variants in patients with ASD.

Interestingly, some of the variants have already been reported before in other affected patients, although they are also present in the general population (Supplementary Table 5). We report a second patient with a previously described rare missense variant in *CNTNAP2 (*p.D1129H*)*. This variant was first described in two monozygotic twins, both affected with ASD and was found to be inherited from their unaffected father (49). Previous functional studies showed that the variant D1129H impairs normal protein traffic, being retained in the endoplasmatic reticulum and inducing proteasomal degradation, probably due to an incorrect protein folding (42). Taking into account the patient described in this study, the allele frequency in ASD cases is 2/1828, whereas it is only found in 1/121410 alleles from the general population. Segregation studies in our case showed that the mutation is inherited from the unaffected father (Figure 2).

Case-control study

Besides the study of highly penetrant variants, we studied the contribution of rare variants following an oligogenic model, by comparing the burden of rare variants between cases and controls. We first compared the load of rare protein-coding variants per gene between patients and controls. This analysis did not yield any significant association when comparing exonic, deleterious (nonsynonymous damaging and LoF variants) or LoF variants, respectively ($p>0.05$ in all cases).

Since ASD present a high degree of genetic heterogeneity, gene-based association studies may be underpowered to detect single locus associations. Therefore, we performed pathway-based association tests to study the contribution of synaptic cell adhesion molecules and scaffold proteins, including 21 functionally related genes (Supplementary Table 1). Comparison of the burden of exonic rare variants of all genes revealed a nearly significant association ($p=0.07$), with rare variants being more prevalent among patients (Table 2). We then carried out separate analysis of LoF intolerant genes (n=10), defined as those with probability of LoF intolerance (pLI) > 0.9 according to ExAC, and LoF tolerant genes (n=11). Interestingly rare variants affecting LoF intolerant genes were significantly more prevalent among cases compared to controls ($p=0.03$), but the same effect was not seen for LoF tolerant genes ($p=0.56$).

	All synaptic genes (n=21)	LoF intolerant synaptic genes (n=10)	LoF tolerant synaptic genes (n=11)
Cases (n= 279)	113/279	63/279	61/279
Controls (n=105)	32/105	12/105	20/105
p.value (Fisher's exact test)	0.07	0.03	0.52

Table 2. Prevalence of rare variants in controls and cases. Fisher's exact test was used to establish statistical significance.

Somatic variation

In addition to rare deleterious SNVs, we also studied the contribution of somatic mosaicism. Taking advantage of our relative high coverage (70x), we studied variants present at lower AB ratios, by applying a more sensitive

variant calling algorithm. Previous studies modeling the statistical probability of detecting mosaicism at differing depths, have shown that at our coverage, we have >90% probability of detecting mosaic events present in approximately 15% of alleles, which corresponds to around 30% of the studied cells (50). In order to discard false positive mosaic events, we selected only variants with an AB ratio <0.2 (which had not been evaluated by our previous approach) and discard previously described variants and those present in more than one individual. After this filtering, only three variants affecting coding regions and with functional consequences were found. All of the variants were discarded after manually examination with IGV, since they corresponded to technical artifacts.

Copy Number Variants (CNVs)

Finally, we analyzed CNVs using XHMM in order to detect smaller exonic CNVs not detected previously by aCGH. After appropriate filtering, we detected a total of 5 unique CNVs. All CNVs except one were found in regions of high variability, with previous variants being reported in the general population. The remaining CNV, a partial exonic deletion affecting exons 4 to 11 of *CNTNAP5* was validated by MLPA. Unfortunately, no parental DNA was available to perform segregation studies.

DISCUSSION

In this study, we have analyzed rare genetic variants affecting a group of candidate genes in a group of controls and ASD patients by targeted sequencing. Our aims were to confirm the role of candidate genes previously identified by exome sequencing, as well as to study the prevalence and extend of rare variants in synaptic cell adhesion and scaffold molecules. ASD present a very complex genetic architecture, with hundreds of affected loci harboring variants of variable penetrance. To identify both type of variation, highly penetrant mutations and variants with milder effects, we have performed validation and segregation studies of probably deleterious variants, as well as gene-based and pathway-based case control studies.

A total of 15 candidate genes with previous evidence from exome sequencing were included in this study (Supplementary Table 1). In six of these genes

(*CDKL5*, *CUL3*, *MED13L*, *KCNV1*, *MAOA* and *SCN2A*), a variant of high penetrance, either *de novo* or X-linked, had been identified. In this study, we have detected additional deleterious variants in *SCN2A* and *CDKL5*, confirming their role in the pathogenesis of ASD. A *de novo* missense variant affecting a highly conserved residue of *CDKL5* (p.Y171C) was identified in a female patient with a phenotype compatible with those previously described (51). The patient presented subtle dysmorphic features, with mid-face hypoplasia, deep-set eyes, a low nasal bridge with anteverted nares and a marked philtrum. In addition, her speech acquisition was delayed with only babble being acquired and she presented hand sterotypies. Unfortunately, no parental DNA was available from the patient carrying the nonsense mutation in *SCN2A*, but since no LoF inherited variants have been described in this gene, it is probably a *de novo* variant. No additional deleterious variants were identified in the rest of the genes (*CUL3*, *MED13L*, *KCNV1*, *MAOA),* reaffirming the high genetic heterogeneity of ASD.

The rest of the genes included to replicate previous exome sequencing findings were genes with inherited variants that could act as susceptibility factors of milder penetrance in an oligogenic model. This group of genes included recurrently mutated genes (*ANK3*, *ASMT*, *NLRP8*, *RBM19* and *SVEP1*) and genes harboring LoF variants in patients but none or very few in EVS (*ALG9*, *DDHD2*, *IGSF10* and *RTN3*). Recent data from collaborative projects integrating large-scale sequencing projects, such as Kaviar and ExAC, has expanded our knowledge about the distribution of LoF variants and gene redundancy in humans. This large amount of data revealed that the probability of LoF intolerance in this group of genes was, in fact, very close to 0. Therefore, since we did not identify any clearly pathogenic additional variant and gene-based case-control studies did not reveal any significant association, we cannot confirm their role in the liability to ASD.

In contrast, our results point towards a prominent role of synaptic cell adhesion molecules and scaffold proteins in the etiology of ASD. First, we identified rare deleterious variants, both *de novo* and inherited segregating with the phenotype. Two *de novo* variants in genes *SHANK3* and *CNTNAP2* were identified. The variant found in *SHANK3* was a stopgain mutation (p.Q1275X), resulting in a premature stop codon and a truncated protein. The patient presented severe intellectual disability, frequent stereotypies, was

non-verbal and had experienced an early regression. In addition, he presented epilepsy with generalized tonic-clonic seizures with an age of onset around 20 months old. His head circumference was <2 SD at age 2. Interestingly, we detected a second LoF mutation in this patient, a maternally inherited nonsense mutation in *CNTN6* (p.Q740X). Mutations in *SHANK3* can result in a varying degree of severity, suggesting that additional genetic variants could act as modifying factors (31). Therefore, the maternal inherited nonsense mutation in *CNTN6* could act as a modifying genetic factor, affecting the degree of severity in this patient. Interestingly, a second-site CNV affecting *CNTN6* was already proposed as two-hit factor in a girl with ASD presenting an inherited 16p11.2 deletion (52). The second *de novo* variant was a missense mutation, (p.S820R) predicted as damaging by various prediction algorithms, in *CNTNAP2.* The patient presenting this mutation showed global developmental delay with marked impairments in language: immediate and delayed echolalia, a stereotyped and repetitive language and pronoun reversal. Interestingly, *CNTNAP2* has been associated to language impairments, being a target of *FOXP2* and affecting language development in the general population (22,53,54).

In addition to *de novo* variants, we have also identified rare inherited variants that could act as susceptibility factors. Two additional pathogenic variants were identified in *CNTNAP2*: a nonsense mutation resulting in a truncating protein and a damaging missense variant previously shown to impair protein traffic. Both variants were inherited from non-affected progenitors, in agreement with previous studies showing a role of both common and rare variation in *CNTNAP2* in the susceptibility to ASD and to related traits (21,23,24,55,56). A novel missense variant in *NRXN2* (p.R728C) was also identified in a family with two affected siblings and their father presenting OCD. Previous studies addressing the overlap between both disorders have shown that the risk of comorbid OCD is increased in individuals with ASD. In addition, individuals with OCD have an increased risk of having children with ASD, suggesting shared genetic liability between both disorders (57,58). Therefore, this variant could act as a genetic risk factor for psychiatric disorders and its phenotypic outcome could depend on additional genetic and environmental factors. The fact that the third unaffected sibling presents SLI suggests that additional variants creating a sensitized background may be present in this family.

Although gene-based case-control association studies of neurexin and neuroligin genes yielded no significant results, we identified a positive association using a pathway-approach. Pathway-based approaches offer significant advantages compared to single locus strategies. They avoid the use of multiple-test correction and are especially well suited for complex diseases presenting with a high degree of genetic heterogeneity, such as ASD, where single loci associations can be difficult to replicate. By comparing the burden of exonic rare variants, we identified a nearly significant increase of rare variation in synaptic cell adhesion and scaffold molecules in ASD patients. However, only LoF intolerant genes were driving the effect of this association, with a significant increase in the burden of rare mutations ($p=0.03$). This result further proves the implication of the synaptic related proteins in the pathogenesis of ASD. Interestingly, nearly all variants included in this analysis have been previously described in the general population, albeit at low frequencies. Therefore, their contribution to the liability to ASD may follow an oligogenic model, in which each variant may increase the risk, but none would be sufficient to cause it.

One of the main limitations of this study is our small sample size. Case-control studies of complex disorders with genetic heterogeneity require large sample sizes since the rarity of the events reduces the statistical power. However, a few approaches can help overcome this limitation. Segregation studies of highly deleterious variants can detect *de novo* and variants cosegregating with the phenotype in families with multiple affected members. In addition, reporting variants can help identify additional cases, such in the case as the previously described variant in *CNTNAP2.* Also, performing functional studies of selected variants can provide further evidence. Finally, pathway-based approaches can help increase the statistical power, as well as to identify altered mechanisms, by combining information from functionally related genes.

By combining all these lines of evidence, we have shown that both *de novo* and rare inherited variants in brain related genes and, specifically in synaptic cell adhesion and scaffold proteins play a role in the pathogenicity of ASD. Our results reinforce the role of previously associated genes, such as *SCN2A* and *CDKL5*, which follow a Mendelian monogenic model, in which *de novo* mutations in a single genetic factor would be sufficient to cause ASD.

Nonetheless, we provide additional cases that point towards a more complex role of *de novo* and inherited variants in the genetic architecture of ASD. For example, we describe both inherited and *de novo* pathogenic variants affecting the same gene, *CNTNAP2*. This suggests that inherited variants may only act in a background of additional risk modifying factors and that some *de novo* variants may need additional risk alleles in order to show their effect. Moreover, inherited variants of milder effect may also act as phenotypic modifying factors that explain the variable range of severity in patients carrying highly penetrant variants, such as the inherited nonsense mutation in *CNTN6* described in a boy with a *de novo* LoF in *SHANK3*. Therefore, this study provides specific examples of the complex genetic architecture of ASD and the role of both *de novo* and inherited variants.

References

1. Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. J Child Psychol Psychiatry [Internet]. 1977 Sep;18(4):297–321. Available from: http://www.ncbi.nlm.nih.gov/pubmed/562353

2. Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM. Concordance for the syndrome of autism in 40 pairs of afflicted twins. Am J Psychiatry [Internet]. 1985 Jan;142(1):74–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/4038442

3. Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, et al. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. J Child Psychol Psychiatry [Internet]. 1989 May;30(3):405–16. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2745591

4. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, et al. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med [Internet]. 1995 Jan [cited 2015 May 16];25(1):63–77. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7792363

5. Taniai H, Nishiyama T, Miyachi T, Imaeda M, Sumi S. Genetic influences on the broad spectrum of autism: study of proband-ascertained twins. Am J Med Genet B Neuropsychiatr Genet [Internet]. 2008 Sep 5;147B(6):844–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18361421

6. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. Arch Pediatr Adolesc Med [Internet]. 2009 Oct;163(10):907–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19805709

7. Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry [Internet]. 2010 Nov;167(11):1357–63. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20686188

8. Ronald A, Hoekstra R a. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156(3):255–74.

9. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics [Internet]. 2011 Sep [cited 2014 Jul 12];128(3):e488–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21844053

10. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA [Internet]. 2014 May 7;311(17):1770–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24794370

11. Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev [Internet]. 2012 Jun;22(3):229–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22463983

12. Ching MSL, Shen Y, Tan W-H, Jeste SS, Morrow EM, Chen X, et al. Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. Am J Med Genet B Neuropsychiatr Genet [Internet]. 2010 Jun 5;153B(4):937–47. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20468056

13. Schaaf CP, Boone PM, Sampath S, Williams C, Bader PI, Mueller JM, et al. Phenotypic spectrum and genotype-phenotype correlations of NRXN1 exon deletions. Eur J Hum Genet [Internet]. 2012 Dec;20(12):1240–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22617343

14. Gauthier J, Siddiqui TJ, Huashan P, Yokomaku D, Hamdan FF, Champagne N, et al. Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. Hum Genet [Internet]. 2011 Oct;130(4):563–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21424692

15. Kim H-G, Kishikawa S, Higgins AW, Seong I-S, Donovan DJ, Shen Y, et al. Disruption of neurexin 1 associated with autism spectrum disorder. Am J Hum Genet [Internet]. 2008 Jan;82(1):199–207. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18179900

16. Vaags AK, Lionel AC, Sato D, Goodenberger M, Stein QP, Curran S, et al. Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am J Hum Genet [Internet]. The American Society of Human Genetics; 2012 Jan 13 [cited 2012 May 30];90(1):133–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22209245

17. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature [Internet]. 2009 May 28;459(7246):569–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19404257

18. Wöhr M, Silverman JL, Scattoni ML, Turner SM, Harris MJ, Saxena R, et al. Developmental delays and reduced pup ultrasonic vocalizations but normal sociability in mice lacking the postsynaptic cell adhesion protein neuroligin2. Behav Brain Res [Internet]. 2013 Aug 15;251:50–64. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/22820233

19. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet [Internet]. 2015 Jun;47(6):582–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25961944

20. Ross JL, Tartaglia N, Merry DE, Dalva M, Zinn AR. Behavioral phenotypes in males with XYY and possible role of increased NLGN4Y expression in autism features. Genes Brain Behav [Internet]. 2015 Feb;14(2):137–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25558953

21. Arking DE, Cutler DJ, Brune CW, Teslovich TM, West K, Ikeda M, et al. A Common Genetic Variant in the Neurexin Superfamily Member CNTNAP2 Increases Familial Risk of Autism. Am J Hum Genet. 2008;82(1):160–4.

22. Whitehouse AJO, Bishop DVM, Ang QW, Pennell CE, Fisher SE. CNTNAP2 variants affect early language development in the general population. Genes Brain Behav [Internet]. 2011 Jun;10(4):451–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21310003

23. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, et al. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet [Internet]. 2008 Jan [cited 2015 Dec 4];82(1):165–73. Available from: /pmc/articles/PMC2253974/?report=abstract

24. Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, Engels H, et al. Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. BMC Med Genet [Internet]. 2011 Jan [cited 2015 Dec 4];12:106. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3162517&tool=pmcentrez&rendertype=abstract

25. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature [Internet]. Nature Publishing Group; 2012 Apr 4 [cited 2012 Apr 4];485(7397):246–50. Available from: http://dx.doi.org/10.1038/nature10989

26. Pagnamenta AT, Bacchelli E, de Jonge M V, Mirza G, Scerri TS, Minopoli F, et al. Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. Biol Psychiatry [Internet]. 2010 Aug 15;68(4):320–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20346443

27. Zuko A, Kleijer KTE, Oguro-Ando A, Kas MJH, van Daalen E, van der Zwaag B, et al. Contactins in the neurobiology of autism. Eur J Pharmacol [Internet]. 2013 Nov 5;719(1-3):63–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23872404

28. Poot M. A candidate gene association study further corroborates involvement of contactin genes in autism. Mol Syndromol [Internet]. 2014 Aug;5(5):229–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25337070

29. Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, et al. Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. Am J Hum Genet [Internet]. 2004 Jun;74(6):1286–93. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15106122

30. Sato D, Lionel AC, Leblond CS, Prasad A, Pinto D, Walker S, et al. SHANK1 Deletions in Males with Autism Spectrum Disorder. Am J Hum Genet [Internet]. 2012 May 4;90(5):879–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22503632

31. Leblond CS, Nava C, Polge A, Gauthier J, Huguet G, Lumbroso S, et al. Meta-analysis of SHANK Mutations in Autism Spectrum Disorders: a gradient of severity in cognitive impairments. PLoS Genet [Internet]. 2014 Sep;10(9):e1004580. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25188300

32. Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, et al. Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. Nat Genet [Internet]. 2010 Jun;42(6):489–91. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20473310

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

34. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. Nat Rev Genet [Internet]. 2014 Feb;15(2):133–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24430941

35. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey a. J, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature [Internet]. 2012 Apr 4 [cited 2012 Apr 3];485(7397):1–6. Available from: http://www.nature.com/doifinder/10.1038/nature10945

36. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, et al. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet. 2011 Jun;43(6):585–9.

37. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature [Internet]. 2014 Nov 13;515(7526):216–21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25363768

38. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature [Internet]. 2012 May 10;485(7397):246–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22495309

39. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature [Internet]. Nature Publishing Group; 2012 Apr 4 [cited 2012 Apr 4];485(7397):242–5. Available from: http://dx.doi.org/10.1038/nature11011

40. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature [Internet]. Nature Publishing Group; 2012 Aug 22 [cited 2012 Aug 22];488(7412):471–5. Available from: http://www.nature.com/doifinder/10.1038/nature11396

41. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron [Internet]. Elsevier Inc.; 2012 Apr 26 [cited 2012 Jul 16];74(2):285–99. Available from: http://dx.doi.org/10.1016/j.neuron.2012.04.009

42. Falivelli G, De Jaco A, Favaloro FL, Kim H, Wilson J, Dubi N, et al. Inherited genetic variants in autism-related CNTNAP2 show perturbed trafficking and ATF6 activation. Hum Mol Genet [Internet]. 2012 Nov 1;21(21):4761–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22872700

43. He X, Sanders SJ, Liu L, De Rubeis S, Lim ET, Sutcliffe JS, et al. Integrated Model of De Novo and Inherited Genetic Variants Yields Greater Power to Identify Risk Genes. Williams SM, editor. PLoS Genet [Internet]. 2013 Aug 15 [cited 2013 Aug 22];9(8):e1003671. Available from: http://dx.plos.org/10.1371/journal.pgen.1003671

44. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature [Internet]. 2014 Oct 29 [cited 2014 Oct 29];515(7526):209–15. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4402723&tool=pmcentrez&rendertype=abstract

45. Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang Y, Casals F, et al. Variation in genome-wide mutation rates within and between human families. Nat Genet [Internet]. 2011 Jul;43(7):712–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21666693

46. Codina-Solà M, Rodríguez-Santiago B, Homs A, Santoyo J, Rigau M, Aznar-Laín G, et al. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. Mol Autism [Internet]. 2015;6:21. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25969726

47. EVS. Exome Variant Server. NHLBI GO Exome Seq Proj [Internet]. 2014; Available from: http://evs.gs.washington.edu/EVS/

48. Glusman G, Caballero J, Mauldin DE, Hood L, Roach JC. Kaviar: an accessible system for testing SNV novelty. Bioinformatics [Internet]. 2011 Nov 15;27(22):3216–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21965822

49. Bakkaloglu B, O'Roak BJ, Louvi A, Gupta AR, Abelson JF, Morgan TM, et al. Molecular cytogenetic analysis and resequencing of contactin associated protein-like 2 in autism spectrum disorders. Am J Hum Genet [Internet]. 2008 Jan;82(1):165–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18179895

50. Acuna-Hidalgo R, Bo T, Kwint MP, van de Vorst M, Pinelli M, Veltman JA, et

al. Post-zygotic Point Mutations Are an Underrecognized Source of De Novo Genomic Variation. Am J Hum Genet [Internet]. 2015 Jul 2;97(1):67–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26054435

51. Fehr S, Wilson M, Downs J, Williams S, Murgia A, Sartori S, et al. The CDKL5 disorder is an independent clinical entity associated with early-onset encephalopathy. Eur J Hum Genet [Internet]. 2013 Mar [cited 2015 Dec 4];21(3):266–73. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3573195&tool=pmcentrez&rendertype=abstract

52. Duyzend MH, Nuttle X, Coe BP, Baker C, Nickerson DA, Bernier R, et al. Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. Am J Hum Genet [Internet]. 2016 Jan 7;98(1):45–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26749307

53. Stromswold K. The genetics of speech and language impairments. N Engl J Med [Internet]. 2008 Nov 27;359(22):2381–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18987364

54. Whalley HC, O'Connell G, Sussmann JE, Peel A, Stanfield AC, Hayiou-Thomas ME, et al. Genetic variation in CNTNAP2 alters brain function during linguistic processing in healthy individuals. Am J Med Genet B Neuropsychiatr Genet [Internet]. 2011 Dec;156B(8):941–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21987501

55. Alarcon M, Abrahams BS, Stone JL, Duvall JA, Perederiy J V., Bomar JM, et al. Linkage , Association , and Gene-Expression Analyses Identify CNTNAP2 as an Autism-Susceptibility Gene. Am J Hum Genet [Internet]. 2008 Jan [cited 2015 Oct 28];82(1):150–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2253955&tool=pmcentrez&rendertype=abstract

56. Stein MB, Yang BZ, Chavira DA, Hitchcock CA, Sung SC, Shipon-Blum E, et al. A common genetic variant in the neurexin superfamily member cntnap2 is associated with increased risk for selective mutism and social anxiety-related traits. Biol Psychiatry. 2011;69(9):825–31.

57. Stewart E, Cancilliere MK, Freeman J, Wellen B, Garcia A, Sapyta J, et al. Elevated Autism Spectrum Disorder Traits in Young Children with OCD. Child Psychiatry Hum Dev [Internet]. 2016 Feb 8; Available from: http://www.ncbi.nlm.nih.gov/pubmed/26858231

58. Meier SM, Petersen L, Schendel DE, Mattheisen M, Mortensen PB, Mors O. Obsessive-Compulsive Disorder and Autism Spectrum Disorders: Longitudinal and Offspring Risk. PLoS One [Internet]. 2015;10(11):e0141703. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26558765

Supplementary Material

Gene	Group	Previous evidence	pLI
ANK3	Codina-Solà et al 2015	Recurrently mutated	1
CDKL5	Codina-Solà et al 2015	de novo variant	1
CNTN1	Contactin		1
CNTN4	Contactin		1
CUL3	Codina-Solà et al 2015	de novo variant	0,97
MED13L	Codina-Solà et al 2015	de novo variant	1
KCNV1	Codina-Solà et al 2015	de novo variant	
MAOA	Codina-Solà et al 2015	X-linked variant	
NLGN2	Neuroligin		0,99
NLGN3	Neuroligin		0,9
NLGN4X	Neuroligin		0,93
NRXN1	Neurexin		1
NRXN2	Neurexin		1
NRXN3	Neurexin		1
SCN2A	Codina-Solà et al 2015	de novo variant	1
SHANK1	Intracelular linking protein		1
SHANK2	Intracelular linking protein		1
SHANK3	Intracelular linking protein		1
ALG9	Codina-Solà et al 2015	LoF variant not present in control database (EVS)	0
ASMT	Codina-Solà et al 2015	Recurrently mutated	0
CNTN2	Contactin		0,13
CNTN3	Contactin		0,01
CNTN4	Contactin		1
CNTN5	Contactin		0
CNTN6	Contactin		0
CNTNAP1	Contactin associated protein		0
CNTNAP2	Contactin associated protein		0
CNTNAP4	Contactin associated protein		0
CNTNAP5	Contactin associated protein		0,1
DDHD2	Codina-Solà et al 2015	LoF variant not present in control database (EVS)	0
DLG2	Intracelular linking protein		0
DLG4	Intracelular linking protein		0,67
IGSF10	Codina-Solà et al 2015	LoF variant not present in control database (EVS)	0
NLGN1	Neuroligin		0,76
NLGN4Y	Neuroligin		0,65
NLRP8	Codina-Solà et al 2015	Recurrently mutated	0
RBM19	Codina-Solà et al 2015	Recurrently mutated	0
RTN3	Codina-Solà et al 2015	LoF variant not present in control database (EVS)	0,04
SVEP1	Codina-Solà et al 2015	Recurrently mutated	0,59

Supplementary Table 1. Genes included and the desing and basis for its inclusion.

Sample	Parental samples	Status	Gene	Exonic function	Functional annotation	Segregation
SS53 11	No	CONTROL	*CNTN2*	Missense	NM_005076:p.S697T	NA
SW201 11	No	CONTROL	*CNTN2*	splicing variant	NM_005076:c.3014-2A>G	NA
GF1293	No	AUT	*SCN2A*	Missense	NM_001040143:p.T365M	NA
GF0991	No	AUT	*SCN2A*	stopgain SNV	NM_001040143:p.R1235X	NA
GF2214	No	AUT	*CUL3*	missense	NM_001257197:p.R62H	NA
GF2243	No	AUT	*CNTN6*	missense	NM_001289081:p.G9V	NA
GF0956	No	AUT	*CNTN6*	missense	NM_001289081:p.G67R	NA
AUT177 01	Yes	AUT	*CNTN6*	stopgain SNV	NM_001289081:p.Q740X	Inherited from non affected mother
AUT302 01	No	AUT	*CNTN3*	missense	NM_020872:p.Y781C	NA
GF1356	No	AUT	*CNTN3*	missense	NM_020872:p.R455G	NA
AUT278 01	Mother only	AUT	*CNTNAP2*	missense	NM_014141:p.G173R	NA
AUT140 01	Yes	AUT	*CNTNAP2*	stopgain SNV	NM_014141:p.W184X	Inherited from non afffected father
AUT192 01	Yes	AUT	*CNTNAP2*	missense	NM_014141:p.S820R	*de novo*
AUT310 01	No	AUT	*CNTNAP2*	missense	NM_014141:p.R938H	NA
SW18 01	No	CONTROL	*ANK3*	missense	NM_020987:p.E2605Q	NA
SW119 11	No	CONTROL	*ANK3*	missense	NM_001149:p.A126V	NA
GF1420	No	AUT	*ANK3*	missense	NM_001204404:p.R788H	NA
GF1416	No	AUT	*ANK3*	missense	NM_001204404:p.R306G	NA
SS26 12	No	CONTROL	*NRXN2*	missense	NM_138734:p.V474F	NA
GF2243	No	AUT	*NRXN2*	missense	NRXN2:p.A1042T	NA

Supplementary Table 2. Loss of function and missense damaging variants observed in our cohort.

Sample	Parental samples	Status	Gene	Exonic function	Functional annotation	Segregation
AUT324 01	Yes	AUT	*NRXN2*	missense	NRXN2:p.R728C	Inherited from father with OCD, also present in affected sibling, but not in not affected sibling
SW196 11	No	CONTROL	*SHANK2*	missense	NM_133266:p.P1156L	NA
SS2 11	No	CONTROL	*ALG9*	missense	NM_001077691:p.L316R	NA
GF1086	No	AUT	*CNTN1*	missense	NM_175038:p.L676M	NA
AUT282 01	Yes	AUT	*RBM19*	frameshift deletion	NM_001146698:c.2196_2203del	Inherited from non affected mother
GF2408	No	AUT	*RBM19*	missense	NM_001146698:p.L9F	NA
AUT207 01	Yes	AUT	*MED13L*	missense	NM_015335:p.R1903S	Inherited from non affected mother
AUT137 01	Yes	AUT	*CNTNAP1*	missense	NM_003632:p.S237R	Inherited from non affected mother
GF1300	No	AUT	*SHANK3*	missense	NM_033517:p.N115Y	NA
GF1808	No	AUT	*SHANK3*	missense	NM_033517:p.G1204R	NA
AUT177 01	Yes	AUT	*SHANK3*	stopgain SNV	NM_033517:p.Q1275X	*de novo*
AUT35 01	Yes	AUT	*CDKL5*	missense	NM_003159:p.Y171C	*de novo*

Supplementary Table 2 (continued)

Gene	Functional annotation	Case AF	Control AF	References (PMID)	Total AF patients	KAVIAR frequency
CNTN4	NM_001206955:p.N178D	1/279	0/105	21308999	2/732	21/154602
NLGN4X	NM_001282146:p.K378R	1/279	0/105	15622415, 1964562	3/1192	1/154602
NRXN3	NM_001105250:p.A7T	1/279	0/105	22209245	2/1258	221/154602
CNTNAP2	NM_014141:p.D1129H	1/279	0/105	18179895, 22872700	2/1828	1/121410
CNTNAP5	NM_130773:p.P714R	2/279	0/105	20346443	3/844	85/154602

Supplementary Table 3. Previously described variants in ASD patients in our cohort.

CHAPTER 3

Genomic and genetic variation at complex segmental duplications in Autism Spectrum disorders

Marta Codina-Solà, Raquel Flores, Cristina Aguado, Aïda Homs, Luis A. Pérez-Jurado, Ivon Cuscó

In preparation

Abstract

Introduction: Autism Spectrum Disorders (ASD) are among the most heritable neurodevelopmental disorders, but their etiology remain unknown in a high proportion of cases. Rearrangements, mutations and conversion events in segmental duplications (SD) largely unstudied could contribute to the missing heritability. To determine their variation and role in ASD, we have studied two candidate regions that harbor recently duplicated genes in the human lineage: the Williams-Beuren syndrome locus at 7q11.23 and the chr9 pericentromeric region.

Methods: We performed targeted capture and sequencing of the coding regions of three multi-copy genes located in SDs (*GTF2I*, *GTF2IRD2 and CNTNAP3)* along with single-copy genes in 216 ASD patients and 64 controls. To identify both Copy Number Variants (CNVs) and Single Nucleotide Variants (SNVs), we mapped the sequences to a single paralogous copy and integrated read depth and relative paralogous sequence quantification. All results were experimentally validated by other methods and extended to a larger case (n=) and control (n=) population.

Results: Copy-gains of 7q11.23 were identified in 2% ASD cases compared to 1% of controls, with duplications of a single SD block (B) significantly increased in patients compared to controls (p=0.02). In addition, we identified a likely pathogenic mutation (p.M1?) in a functional copy of *GTF2IRD2* in two affected brothers, inherited from their father. At *CNTNAP3*, a *de novo deletion* of *CNTNAP3B* was identified in a patient but the global distribution of rearrangements and rare SNVs did not differ significantly between groups.

Conclusions: Our approach successfully identified rearrangements and SNVs in highly identical SDs, including copy-gains and mutations at *GTF2IRD2* or deletions of *CNTNAP3* that could represent susceptibility factors for ASD.

Introduction

Autism Spectrum Disorders (ASD) are a group of neurodevelopmental disorders of early onset, characterized by impairments in social communication and interaction and the presence of repetitive and restrictive behaviours. They are among the most heritable of neurodevelopmental disorders, but currently a genetic cause can be identified in only 20 to 33% of cases (1). Part of the missing heritability of ASD and other complex diseases could lye in unexplored variation in regions of segmental duplications (SD) (2–4). SD are fragments of DNA that range from 1 to 400kb, are present in at least two locations in the genome and share a high degree of identity (>95%). They comprise approximately 5% of the human genome and are among the most dynamic regions of the genome (5,6). Due to their high sequence identity, they are substrates for nonallelic homologous recombination (NAHR) and predispose to several genomic disorders, as well as to polymorphic variation in the population. Some genes contained in SD are likely to play an important role in primate and human evolution, since they show strong signatures of positive selection and are enriched in functional categories related to environment adaptation (7). Therefore, they are excellent candidates to contribute to the interspecies phenotypic variability related to human specific traits, such as language or social behavior.

Despite their possible role in disease and evolution, variation in multi-copy regions remains largely unexplored due to their inherent complexity. Earlier studies using aCGH pinpointed specific regions and demonstrated its high variability, but were limited to duplications in the lower range of copies (8,9). Recently, several studies have used NGS data to infer the range in copy number of multiallelic CNVs (mCNVs) and even to precisely define integer copy number states by taking advantage of the quantitative nature of short-read mapping (10,11). Sequencing approaches offer additional advantages by providing qualitative and quantitative information of the underlying sequence. Quantification of Paralogous Sequence Variants (PSVs) has been used to identify the specific rearranged copy, as well as the specific breakpoints inside SD (10,12,13). Although genome-wide studies offer insights into the extend and the variability of SD, studies targeting specific regions can apply customized computational and molecular strategies to precisely identify and validate rearrangements (14–16).

In this study, we focused on two such complex regions: the 7q11.23 locus, involved in Williams Beuren syndrome (WBS) and its reciprocal duplication syndrome, and the pericentromeric region of chromosome 9. Both have evolved rapidly during the primate lineage, are uniquely duplicated in the human species and contain genes with high brain expression. Therefore, they may play a role in human adaptative behavior and, ultimately, in neurodevelopmental disorders.

The 7q11.23 region is composed of a single copy interval flanked by three large and complex SD located in the centromeric (c), medial (m) and telomeric (t) sides. Each SD is composed of three blocks with different order and orientation, named A, B and C (Figure 1). The region appears to be highly polymorphic in humans and prone to different rearrangements, since paracentric inversions and variation in block copy number have been described in the general population (17,18). This region contains two genes related to cognition, *GTF2I* and *GTF2IRD2*. *GTF2I* has been postulated as the main candidate for the neurocognitive profile observed in WBS and *GTF2IRD2* as a possible modulator (19–22). *GTF2I* has a single functional copy located in the Bm block and two pseudogenes in the Bc and Bt block, whereas *GTF2IRD2* has two functional copies located in the Bm and Bt blocks and one pseudogene in the Bc, that lacks the first two exons (23,24). Interestingly, this region has evolved rapidly in the hominoid lineage and has species-specific rearrangements, such as the duplication of the B block, that is present only in the human lineage (10,25). Therefore, humans are the only species with two functional copies of *GTF2IRD2*.

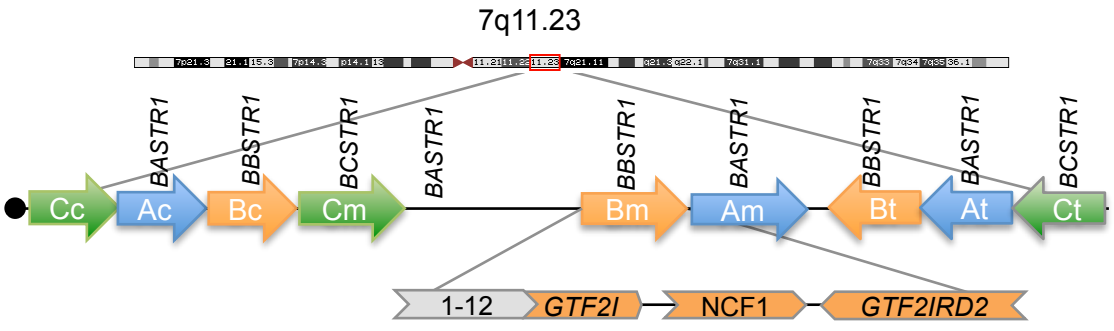


Figure 1. Schematic representation of the 7q11.23 region. The centromeric (c), telomeric (t) and medial (m) SD are showed in arrows with their blocks colored in blue (A), orange (B) and green (C). A magnified image of medial block B shows the orientation of genes and the only functional copy of *GTF2I*. The location of the block-specific short-tandem repeats (BASTR1, BBSTR1, BCSTR1) evaluated is shown.

The pericentromeric region of chromosome 9 is very complex with multiple copies of SDs of high sequence identity (>95%) at both edges of the heterochromatic regions flanking the centromere. It is one of the most dynamic regions of the human genome and, like the 7q11.23 locus, it contains a gene specifically duplicated in the human lineage: *CNTNAP3* (26–28). *CNTNAP3* belongs to the neurexin family of genes, which are cell recognition molecules that ligate to membrane-bound neuroliguins to mediate neuronal interactions. Both neurexins and neuroliguins have been extensively associated to ASD (29). *CNTNAP3* is highly expressed in human brain and has been suggested to mediate neuron-glial interactions in vertebrates and be involved in vesicle transport along the axon (30). According to the hg19 reference genome, humans have two complete functional copies of *CNTNAP3*, located in the 9p13.1 and 9p11.2 chromosome bands termed *CNTNAP3A* and *CNTNAP3B*, respectively and seven partial pseudogenes, that share around 99% of the sequence with the two functional copies (Figure 2). Both *CNTNAP3A* and *CNTNAP3B* have 24 exons that can give rise to multiple isoforms. Due to its complexity and variability, the region has changed substantially on the latest human reference genome assemblies.

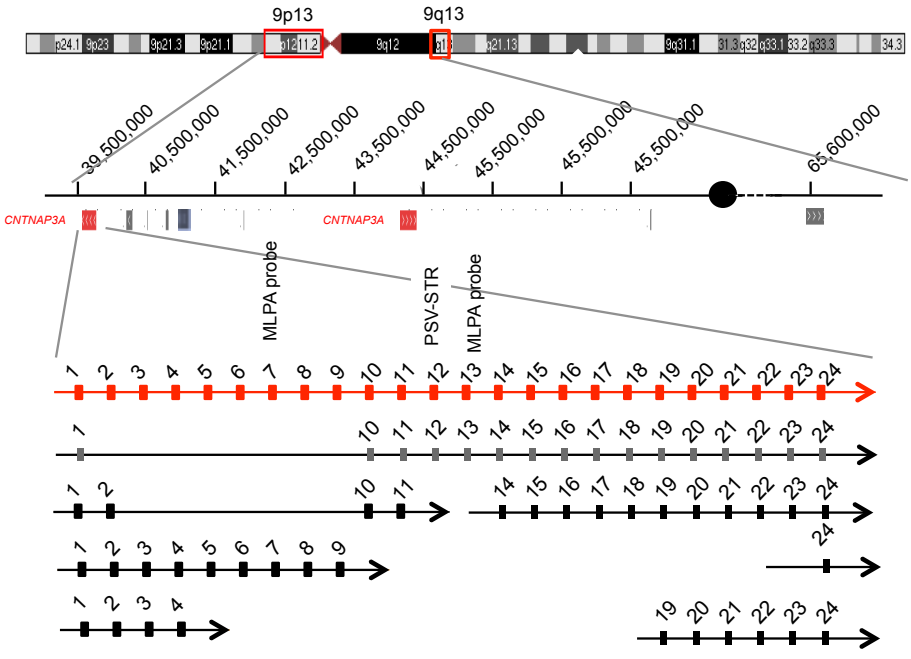


Figure 2. Schematic representation of the pericentromeric region of chromosome 9, with functional copies *CNTNAP3A* and *CNTNAP3B* depicted in red and partial non-functional copies in grey, according to the hg19 human reference genome assembly. The

relative orientation of the copies is shown by the direction of the arrows. Below, the enlarged figure shows the extend of partial copies of *CNTNAP3A* and *CNTNAP3B*, showing the exon content of each one.

To determine the variation and explore the role of these two candidate hot-spot regions in ASD, we have performed targeted sequencing of candidate multi-copy genes *GTF2I*, *GTF2IRD2* and *CNTNAP3* in 216 ASD patients and 64 controls. To identify CNVs, SNVs and conversion events, we have applied a customized computational pipeline, mapping sequences to a single paralogous copy and integrating read depth and relative quantification of PSVs. All results have been experimentally validated and extended in an additional case and control cohort.

MATERIAL AND METHODS

Subjects

We studied 216 European unrelated patients with a diagnosis of idiopathic ASD. The diagnosis was based on the Diagnosis and Statistical Manual of Mental Diseases IV (DSM-IV). The study was approved by the Clinical Research Ethics Committee of the involved centers (CEIC-Parc Salut Mar), and informed consent for participation was obtained from the parents or legal caregivers. Parental and familial samples were obtained from the available relatives who gave informed consent. Results were replicated in an additional group of patients from the same ancestry (n=186).

Regarding control samples, we analyzed 64 unrelated individuals from a cohort of Spanish anonymous blood donors. Genomic DNA was extracted from peripheral blood by the salting-out method using a Puregene DNA purification Kit (Gentra Systems). Results were replicated in a group of additional controls (n=827).

Library preparation, capture and sequencing

DNA was fragmented using a Covaris ultrasonicator (Covaris, Woburn, USA) to 200-400 bp fragments. Libraries were prepared following Illumina's TruSeq standard protocol (Illumina, Inc., San Diego, CA, USA) and double-indexed samples were pooled together. Target genes (*GTF2I, GTF2IRD2* and *CNTNAP3)* and 60 additional single copy genes included as control target,

were captured using a custom design panel by NimbleGen SeqCap EZ custom Library (Hofman-La Roche, Basel, Switzerland). Finally, enriched fragments were paired-end sequenced at a read length of 150 bp on an Illumina MiSeq (Illumina, Inc., San Diego, CA, USA).

Quality control, adaptor trimming and mapping

Quality control was assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adaptor sequences were trimmed using Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Sequences were mapped to a single paralogous copy to avoid arbitrarily reporting of only one of the possible aligning positions produced by most mapping algorithms. This strategy avoids dispersion of the read depth signal to all copies and increases the power to detect small changes in relative depth of coverage (DOC) and variants with lower AB ratios, compatible with heterozygous mutations in multicopy regions. To this end, we constructed an artificial genome reference in which all but one functional paralogous copy of each duplicated genes were masked to N (Bm-*GTF2I* and Bm-*GTF2IRD2* and *CNTNAP3A*) using Bedtools utility *maskfasta*. Genomic coordinates of this genome construct were based on the hg19 version of the human genome. Mapping was performed with BWA using standard parameters, which consider a 2% uniform base error rate and a default maximum edit distance of 0.04, allowing up to 5 mismatches for a read of 150 bp of length. This strategy resulted in a single location per each read that could correspond to any of the paralogous copies of each gene.

CNV detection

Standard CNV detection algorithms are usually designed for the analysis of variation in single copy regions. Therefore, they are not appropriate for detection of changes in multi-copy sites, since its signal decreases as the underlying number of duplicated segments increases. We developed a pipeline to analyze CNVs in SD, which is based on the assumption that relative DOC is stable among samples. First, read depth was computed for probe, for gene and for region using GATK's DepthOfCoverage function. To reduce background signal, targets with low coverage (<40x) were removed.

Finally, normalization was performed against 60 single copy control target regions depth, per sample's per-target, per-gene and per-region.

CNVs in 7q11.23 region, had been previously reported to be rare (1% of general population) (17). Due to its rarity, they are not expected to affect the distribution of relative coverage, which follows a Gaussian distribution. Therefore, to detect CNVs, we compared and plotted each sample's relative coverage to the baseline distribution calculated by the mean of all samples and those individuals with values deviating from the expected normal distribution were selected for validation.

In contrast, the pericentromeric 9 region has been reported to be highly polymorphic, affecting the relative coverage distribution (11). Therefore we followed a different method to accurately determine the integer copy number state of *CNTNAP3A*, *CNTNAP3B* and its partial copies. We calculated standardized RPKM z-scores (ZRPKM) per probe. ZRPKM were analyzed in two different windows according the overlap with the two largest partial copies (exons 1-9 and exons 10-24). Values were plotted in order to establish thresholds to assign absolute copy number states (Supplementary Figure 1).

PSV analyses

To characterize the identified rearrangements, we studied Paralogous Sequence Variants (PSVs). First, we performed a multiple alignment of the reference sequence of duplicated genes and annotated the positions differing between copies. In the 7q11.23 region, we identified a total of 7 PSVs in *GTF2I* and 14 in *GTF2IRD2* located in exons or nearby positions (+/- 150bp). Regarding chromosome 9, we identified 27 PSVs. Previous studies from the Genotype-Tissue Expression project (GTEx) have shown that, whereas *CNTNAP3A* is expressed both in blood and brain (among other tissues), the expression of *CNTNAP3B* is restricted to brain (31). To precisely assign to the PSV with the functional copies *(CNTNAP3A* and *CNTNAP3B)* we performed an expression screening in 40 samples from Peripheral Blood Mononuclear Cells (PBMCs) and three samples from fetal brain tissue by retro-transcriptase PCR and Sanger sequencing.

To compute PSV ratios, we extracted the number of reads with each position and calculated the ratio between the reads belonging to the functional copy to

the total coverage (Supplementary Figure 2). Only positions covered with at least 50 reads were considered and PSVs that showed a high level of variance, consistent polymorphic variation, were discarded. To assign the observed CNVs to the specific paralogous copy we compared the observed ratios to the expected for each rearrangement. Rearrangements were assigned to the best fit model.

SNV analyses

Since SNVs in multi-copy regions are expected to have a lower allelic balance (AB) compared to those in single copy regions, we adapted variant calling and filtering to allow their detection. Variant calling using GATK and standard quality filtering parameters were applied. SNVs and indel calls were only considered if positions had a depth of coverage of at least 10x. Variants with AB ratios lower than 0.2 (the standard AB ratio) were kept and annotated using ANNOVAR, taking into account frequency in control databases (dbSNP142 and Exome Aggregation Consortium [EXAC] database). The nature of the changes was assessed by PolyPhen and Condel protein effect prediction algorithms(http://bg.upf.edu/fannsdb/). Variants present in more than one individual were excluded, since they could represent PSVs or polymorphic variants not previously described in control databases due to the complexity of the regions. PSVs located within read distance (+/- 150bp) were annotated and variants were manually inspected using Integrative Genomics Viewer (IGV) to examine if both changes were in *cis* and could be assigned to a specific copy.

Genomic rearrangements validation

Genomic rearrangements were validated by various techniques: relative quantification of short tandem repeats (STRs) and site-specific nucleotides (SSNs) and Multiple Ligation Probe Amplification (MLPA).

For the 7q11.23 region, STRs located on block A (D7S489 or BASTR1), block B (BBSTR1 and BBSTR2), and block C (POM121) of the LCRs flanking the WBS locus were amplified in order to confirm rearrangements of B blocks and identify the number of copies of A and C blocks, following conditions previously reported in (17,32). Quantification of the number of alleles at multiloci microsatellites was performed by comparing the relative ratios of the area under the peaks from alleles of the same size in different samples, using

the GeneScan 3.1 software (PE Applied Biosystems) as previously described (32). To validate rearrangements affecting the B block and to measure the number of functional copies of *NCF1* we analyzed ratios of a 2-bp deletion PSV located in *NCF1*. This PSV differs in size between the Bc/Bt blocks corresponding to *NCF1* pseudogenes (99 bp) and the Bm block containing the functional copy (101 bp) (33,34). Finally, to quantify the number of functional copies of *GTF2IRD2*, we analyzed a site-specific PSV (SSN11a) located at the 5' part of the gene (intron 2-3), which is only present in the two functional copies located at Bm and Bt blocks. This site-specific PSV corresponds to a single nucleotide change detected using the restriction enzyme BtnII (New England Biolabs), followed by size fractionation on a 3% MetaPhor agarose gel (32). To quantify the relative number of copies of each type of block, we measured the intensity of bands corresponding to the Bm and Bt blocks using the ImageJ software (35). Relative intensities were calculated by means of a dosage quotient between the Bt and Bm blocks, that can be distinguished relative to the others, using several controls as reference values in the same experiment. Finally, the number of functional copies of *GTF2I* was measured by MLPA with specific custom probes targeting the single-copy region. MLPA reactions were carried out under standard conditions and the relative peak height method was used to determine the copy number status.

Regarding the pericentromeric region of chromosome 9, rearrangements were validated by two main techniques. We designed a custom MLPA panel with specific probes to quantify PSVs belonging to *CNTNAP3A* and *CNTNAP3B* as well as the total number of copies (Supplementary Table 1). MLPA was previously shown to be a robust method to quantify the number of copies in mCNVs (36). MLPA reactions were carried out under standard conditions and the relative peak height method was used to determine the copy number status.

Rearrangements were required to meet cut-off values in the total and in one specific copy quantification, which were set to ≤0.8 for deletions and ≥1.1 for duplications.To confirm the identified rearrangements and to perform segregation studies on available parental samples, we designed a second assay based on a PSV-STR located in intron 12 that differs in length between the *CNTNAP3 (*342 bp) *CNTNAP3B (*338 bp) and the partial copy located at

65 Mb that contains exons 10-24 (328 bp) (Supplementary Table 1). PCR reactions were carried under standard conditions and the relative number of alleles was quantified by comparing the relative height of the peaks, measured using the GeneScan 3.1 software (PE Applied Biosystems).

SNVs validation

To perform validation and segregation studies in patients and available relatives of point mutations and small indels, we used PCR followed by Sanger sequencing by capillary electrophoresis (ABI PRISM 7900HT, Applied Biosystems, Foster City, CA, USA). In addition, we performed reverse transcriptase PCR (RT-PCR) from blood mRNA in patients with available samples, followed by standard PCR and Sanger sequencing to determine if the SNVs were present in functional expressed copies.

Statistical analysis

Allele frequencies in ASD patients and control individuals were compared using Pearson's $\chi 2$ and Fisher's exact tests. Statistical analysis was performed using R-platform software. A p-value < 0.05 denoted the presence of statistically significant differences.

RESULTS

Validation of known rearrangements

To confirm that our method can successfully identify CNVs in duplicated regions, we included samples with known rearrangements as positive controls in the design. For the 7q11.23 region, we included four WBS patients with the four possible types of B recombinant blocks already described and two control individuals carrying a confirmed deletion or a duplication only of the duplicated blocks (33). The four WBS patients showed a relative decrease in depth of coverage with respect to the rest of individuals ($\bar{X}$=0.47, SD=0.12) in the single copy region of *GTF2I* (exons 1-12), compatible with a hemizygous deletion. The combined relative depth of coverage for regions of *GTF2I* and *GTF2IRD2* contained in SDs was also clearly decreased with respect to the rest of individuals ($\bar{X}$ =0.77, SD=0.08), (Figure 3A). Rearrangements in the control samples were also validated, with a relative

depth of coverage with respect to controls of 0.88 and 1.14 for the deletion and duplication respectively. Both values were compatible with the expected values: 0.83 (5/6) for the deletion and 1.16 (7/6) for the duplication.

To further characterize the type of rearrangement, we analyzed PSVs in *GTF2IRD2* and calculated the PSV ratio with respect to the medial copy by dividing the number of reads with PSVs belonging to the medial copy to the total number of reads. This operation can result in three different ratios, depending on the specificity of the PSV to tag paralogous copies: Bm/Total, Bm+Bc/Total or Bm+Bt/Total. Depending on the deletion breakpoint, patients with WBS can present four types of recombinant B blocks, that result in different number and type of paralogous copies of *GTF2IRD2* (33). If the deletion is mediated by NAHR between A blocks, both the Bc and Bm blocks would be lost, resulting in four copies of *GTF2IRD2* (1C+1M+2T). If recombination occurs before or within *GTF2IRD2*, it would result in five copies of *GTF2IRD2*, either 1C+2M+2T in the first case and 1C+2T and one chimeric copy between the medial and telomeric chimeric blocks in the second (1C+2T+2Q). Finally, a deletion mediated by inversion, would result in the loss of a medial copy and a gain of a telomeric copy of *GTF2IRD2*, resulting in a 1C+1M+3T genotype, as represented in Figure 3C. Following this method, we were able to correctly assign each patient to its correct recombinant B block type. PSV ratios from the patient bearing the 1C+1M+3T recombinant block B are shown in Figure 3B. Rearrangements in controls were also correctly assigned to the centromeric block, as previously described (17).

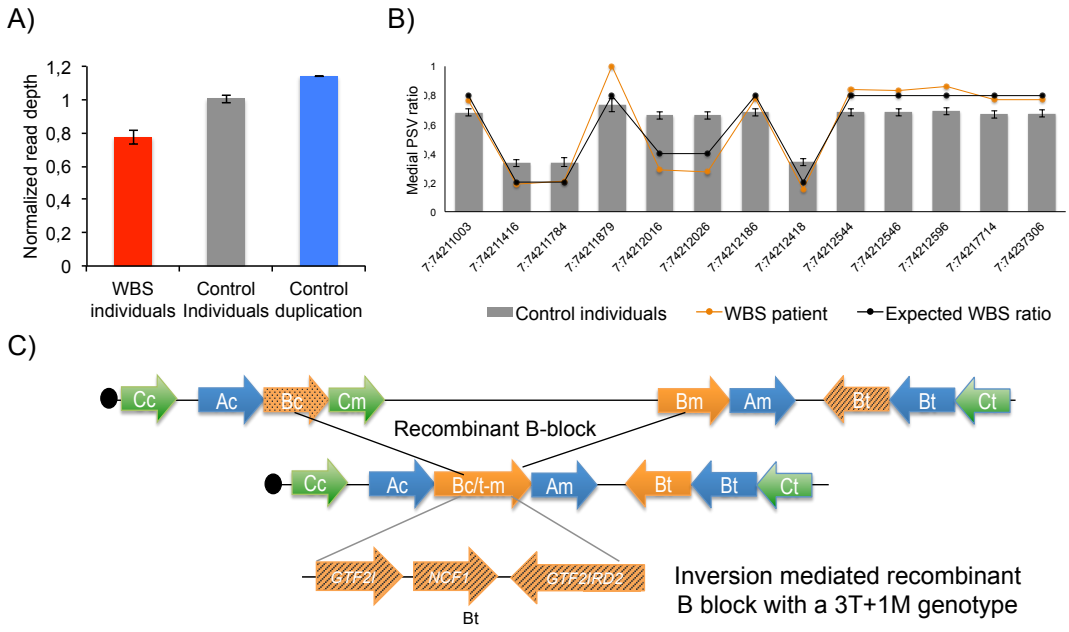


Figure 3. Validation of known rearrangements in individuals with WBS. A) Relative depth of coverage in regions of *GTF2I* and *GTF2IRD2* is decreased in individuals with WBS compared to controls ($\bar{X}$=0.77, SD=0.08). This decrease is similar to the expected ratio theoretical ratio ($5/6 \approx 0.83$). B) The PSV ratio of a WBS patient with a 1C+1M+3T genotype shows values closer to the expected ratio than to that of control individuals. C) Schematic representation of the common deletion associated with WBS.

Regarding *CNTNAP3,* we included three patients with ASD who presented rearrangements previously detected by comparative genomic hybridization (aCGH) with a BAC array: two duplications (CDup1, CDup2) and one deletion (CDel). All rearrangements were experimentally validated by a short tandem repeat (STR-PSV) located in intron 12, that varies in length between the copies predicted to contain this region (*CNTNAP3A*, *CNTNAP3B* and the largest partial copy located in 65 Mb) (Supplementary Table 1). This assay allows the relative quantification of copies, as well as their assignment to a specific paralogous copy, although its specificity is limited by the expected high degree of conversion between copies. By calculating a dosage quotient between alleles, we were able to determine that the two individuals with complete duplications and the individual bearing a deletion carried 7 and 5 alleles respectively (Figure 4B).

Our method to detect CNVs in multi-copy regions based on the relative depth of coverage, confirmed all rearrangements, since all individuals showed

values compatible with each specific rearrangement, as shown in Figure 4A. Patients bearing duplications showed an increment in relative depth of coverage of 1,25 (CDup1) and 1.25 (CDup2), compared to 0.69 in the patient carrying a deletion (Figure 4A).

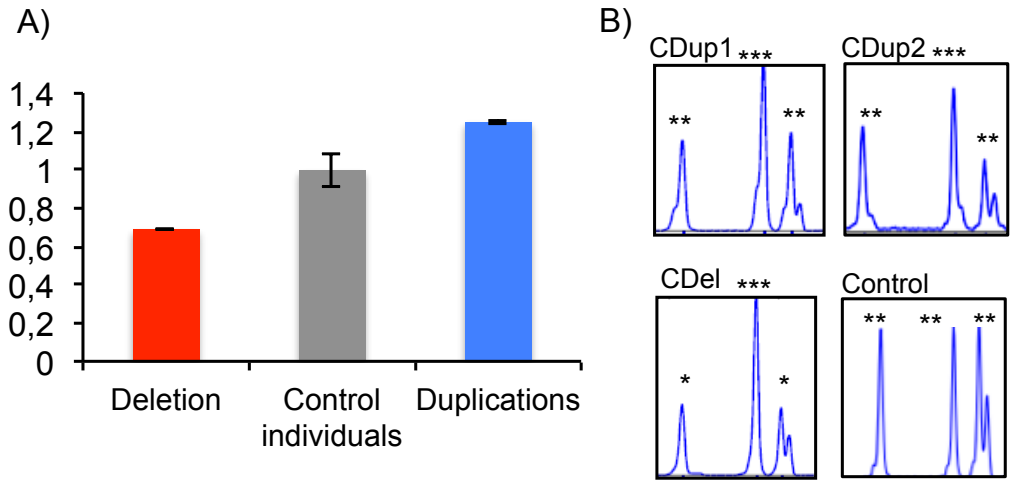


Figure 4. Detection of previously experimentally validated rearrangements in the pericentromeric region of chr9. **A)** Relative depth of coverage analysis showed an increase in individuals with validated duplications and a decrease in the individual carrying a deletion. **B)** Experimental validation of rearrangements by a PSV-STR located in intron 12 of *CNTNAP3*. The peaks (from left to right) correspond to the partial copy located in 65Mb, *CNTNAP3B* and *CNTNAP3A* respectively.

CNVs at 7q11.23

After applying our CNV detection method to the genes contained in the region (*GTF2I* and *GTF2IRD2*), we identified a total of 6/216 (2,9%) of patients bearing a CNV, compared to 1/64 (1,56%) of controls. Of these, 2/216 patients (0,95%) presented a deletion compared to 1/64 controls (1,6%). Duplications of B blocks were found in 4/216 patients (1,85%), but none was found in controls (Table 1). All rearrangements were experimentally validated by means of a previously described PSV located in exon 2 of *NCF1 (NCF1*-PSV*)* (32,33). Segregation studies in all patients with available parental samples showed that all rearrangements were inherited from unaffected progenitors (Supplementary Table 2).

Since B block duplications were more frequent in patients than in our cohort of controls and in previously described data (17), we extended the analysis in

additional cohorts by genotyping of the *NCF1*-PSV. We detected 4/186 (2,1%) duplications and 2/186 (1,1%) deletions. In the extended cohort of controls, we detected 10/827 (1,1%) duplications and 9/827 (1,0%) deletions. Overall, the frequency of duplications was two times higher (2%, 8/395) in cases than in controls (1.08%, 10/929) (Table 1), but did not reach statistical significance (p=0.19, Fisher's exact test).

	Read depth		*NCF1*-PSV		Total	
	Cases	Controls	Cases	Controls	Cases	Controls
Del	2/216	1/64	2/186	9/827	4/395 (1.0%)	10/929 (1.1%)
Dup	4/216	0/64	4/186	10/827	8/395 (2.0%)	10/929 (1.1%)
All	6/216	1/64	6/186	19/827	11/395 (2.8%)	20/929 (2.1%)

Table 1. Frequency of gains and losses of B blocks among cases and controls detected by two methods: targeted sequencing and analysis of the PSV located at *NCF1*.

Molecular characterization of CNVs in 7q11.23

To further characterize the rearrangements, we analyzed various microsatellites and PSVs located within the duplicated blocks. First, we defined the extend of the rearrangements by relative quantification of two microsatellelites located at the adjacent A and and C blocks (BBSTRA and POM121). Three types of rearrangements had been previously defined: WBS-CNV1 (affecting A, B and C blocks), WBS-CNV2 (including A and B blocks) and WBS-CNV3 (containing only blocks B) (17). Regarding deletions, no significant differences in the type of rearrangements were observed between cases and controls (Table 2). In contrast, one of the duplication types (WBS-CNV3dup) was significantly more prevalent in cases than controls (p=0.02, Fisher's exact test) (Table 2).

	Deletions		Duplications	
	Cases	Controls	Cases	Controls
WBS-CNV1	3	7	1	5
WBS-CNV2	0	3	1	2
WBS-CNV3	1	0	6	3
Total	4/395	10/929	8/395	10/929

Table 2. Type of rearrangements identified in cases and controls according to the extend of the CNVs.

We next defined the origin of the specific rearranged block and the functional genes affected in the rearrangements, by the analysis of various PSVs and microsatellite loci. The B block contains three genes: *GTF2I, GTF2IRD2,* and *NCF1*. Of the three copies of *GTF2I*, only the one located in the Bm block is functional, since the rest of the copies lack the 5' region of the gene. Since none of the individuals with rearrangements showed changes in the single-copy region of this gene by MLPA, none of the rearrangements affect a functional copy of this gene. Regarding *NCF1*, its number of functional copies can be easily quantified by measuring the already genotyped indel PSV located in exon 2 (Supplementary Table 2).

GTF2IRD2 has two functional copies located in Bm and Bt blocks, whereas the centromeric copy lacks the two first exons and is not transcribed. To define the origin of the rearranged copies of *GTF2IRD2* in those individuals with available sequencing data, we studied PSV-ratios (Figure 5, Supplementary Table 3 and 4). For the rest of individuals we analyzed a site-specific PSV (SSN11a), which is located at intron 3 of *GTF2IRD2.* Since this intron is present in the 5' part of the gene, it is contained in the telomeric and medial copies, but not in the centromeric copy. Regarding deletions, all rearrangements except one were of centromeric origin, with agreement between results from SSN11a and PSV ratios (Figure 5, Supplementary Table 2 and 3). The remaining deletion was of telomeric origin, hence encompassing a functional copy of *GTF2IRD2*, and was found in a patient (Supplementary Table 2).

As for duplications, results from SSN11a at intron 3 showed that two of the gains were of medial origin, containing a functional copy of *GTF2IRD2*. They were found in a patient and a control and both were of the WBS-CNV1 type (Supplementary Table 2). According to SSN11a, located in intron 3, the rest of the duplications were of centromeric origin, since no gain was detected at this point. However, the detailed characterization by PSV-ratios in four cases with available sequencing data showed compatible results with a gain of a telomeric or medial copy (Figure 5, Supplementary Table 4). All four cases had a WBS-CNV3 type duplication, which was previously suggested to be of telomeric or medial origin (17). Interestingly, the frequency of WBS-CNV3 duplications was significantly increased in cases compared to controls. Taken together, the results from PSV-ratios and SSN11a suggest that duplications

in these individuals encompass a partial telomeric or medial functional block, with its breakpoint located between exon 5 and intron 3 of *GTF2IRD2* (Figure 5). The inability to distinguish medial and telemoreic rearrangements is due to the high degree of identity between the two copies, resulting in few specific PSVs. In contrast, rearrangements affecting centromeric copies can be more easily identified, as shown by the correct assignment of a positive control included in the design (C2) (Figure 5, Supplementary Table 4).

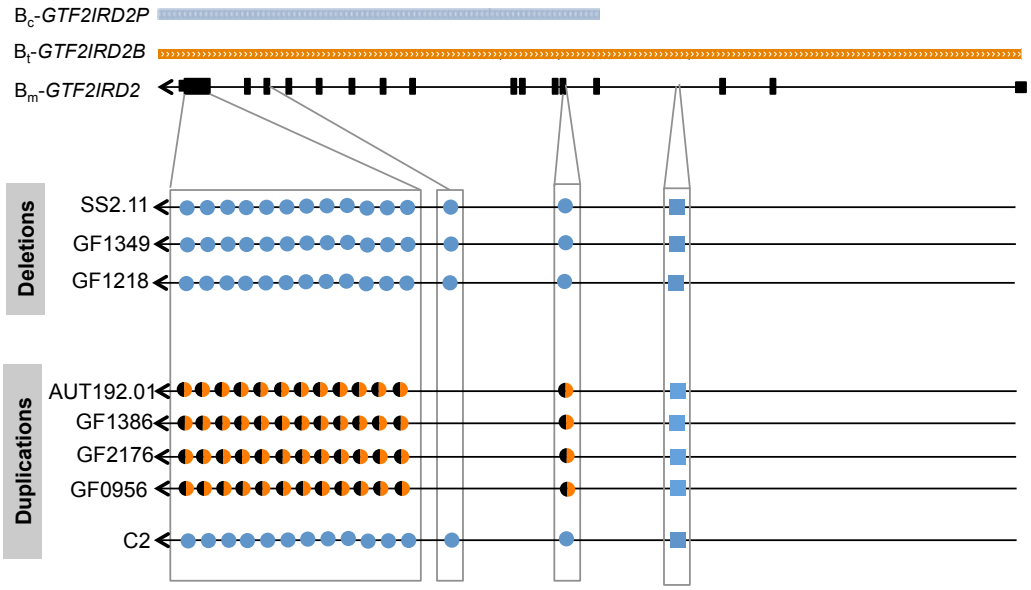


Figure 5. Schematic representation of the detailed genetic architecture and PSV content of *GTF2IRD2* and its paralogous copies. Two functional copies encompassing the entire gene are located at the medial (black) and telomeric (orange) blocks, whereas the non functional copy lacking exons 1 and 2 is located in the centromeric B block (blue). The location of different PSVs is shown below and are colored according to its origin. Circles showed origin established by analysis of PSV-ratios obtained from targeted sequencing, whereas squares refer to information obtained by SSN experimental validation.

Rare SNVs affecting coding regions of *GTF2I* and *GTF2IRD2*

Besides genomic rearrangements, pathogenic SNVs in multi-copy genes could also act as susceptibility factors to ASD. Therefore, we studied SNVs in the coding regions of *GTF2I* and *GTF2IRD2*. To exclude PSVs and SNPs not previously described in control databases, we filtered variants present in more than one individual. The nature of the changes was assessed by

various protein effect prediction algorithms, with respect to the medial copy of both genes.

In *GTF2I*, we detected a total of 16 SNVs: 10 SNVs in patients (n=279) and 6 SNVs in controls (n=105) (Supplementary Table 5). All SNVs presented AB ratios consistent with a heterozygous state in multi-copy regions, which is expected to be around 16% (mean AB ratio=16%, range=9-28%). To assign each SNV to a specific copy, we annotated PSVs in read distance length and manually examined if both variants were present in *cis* or performed experimental approaches, such as long-range PCR and sequencing. Among patients, we detected only one variant that could be clearly pathogenic: a 1bp insertion in exon 13 of *GTF2I*, that was predicted to give rise to a stop codon shortly after. To determine if it affected a functional copy of *GTF2I*, we performed long range PCR with primers located at exon 12 and exon 13, followed by Sanger sequencing. The medial functional copy of *GTF2I* is the only containing exons 1 to 12, and, therefore, the only amplified in this assay. Since the variant was not detected, it must be located in the non-functional telomeric or centromeric copies. The rest of the variants in patients were missense. To determine their possible contribution, we validated and performed segregation studies on all individuals with available parental samples. All SNVs were validated and all were inherited from a non-affected progenitor (Supplementary Table 5).

Regarding *GTF2IRD2*, only one rare SNV was identified in a single patient, that was predicted to affect the start codon (c.T2G, p.M1?). Since this SNV is located on the first exon of *GTF2IRD2*, it must affect one of the two functional copies, either the telomeric or medial copy. Validation and segregation studies by Sanger Sequencing revealed that the mutation was also present in an affected brother and was inherited from his unaffected father. Sequencing of blood mRNA of the three carriers confirmed that the mutation is located in a functional expressed copy. In *silico* prediction of the resulting possible open reading frames (ORF) showed that there is a nearby in-frame alternative start codon at position +21 (Met21), that would result in a protein lacking the first 20 aminoacids.

Genetic architecture of the pericentromeric region of chr9

In order to identify rearrangements in this complex region, we have analyzed the total copy number integer state of *CNTNAP3A*, *CNTNAP3B* and its partial copies. Since most variation probably affects pseudogenes and may occur independently in different paralogous copies, we have determined copy number states in two window regions, defined by its overlap with the two main partial copies. The results show that the region is highly polymorphic, especially for exons 2-9, that are contained in the partial copy located at 9p (Table 3). The two window regions overlapping with the largest partial copies showed different frequencies and discordant copy number states in the same individual, suggesting that the two partial copies are responsible for most variation and are segregating independently.

		Integer copy number state			
		4	**5**	**6**	**7**
Exons 2-9	**Cases (n=216)**	14 (6%)	65 (30%)	118 (55%)	19 (9%)
	Controls (n=64)	3 (5%)	19 (29%)	36 (56%)	7 (11%)
Exons 10-24	**Cases (n=216)**	0 (0%)	11 (5%)	202 (94%)	3 (1%)
	Controls (n=64)	0 (0%)	3 (5%)	56 (87%)	5 (8%)

Table 3. Inferred copy number states from ZPRKM data for exons overlapping with the two main partial copies.

To confirm the validity of our method, we compared our distribution of integer copy number calls for exons 2-9 to that recently reported for European population in a genome-wide study evaluating multiallelic CNVs (11). The results showed a similar CNV distribution, with 30% of our cohort carrying a heterozygous deletion of the partial copy containing exons 2-9, a very similar proportion to that previously described. The similar distribution in both cohorts confirms that our method can correctly assign total integer copy number states (Figure 6A). No significant differences were found between patients and controls (Figure 6B).

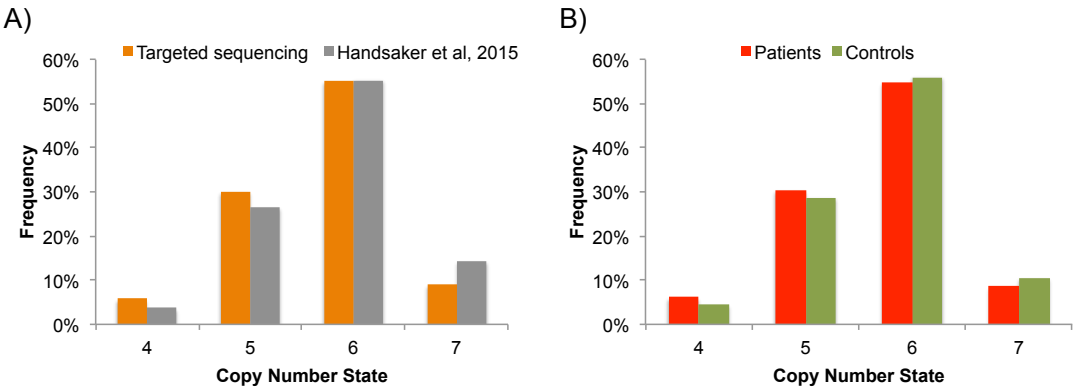


Figure 6. A) Copy number state inferred from ZRPKM distribution in our cohort (orange) compared to that described in Handsaker et al. (grey) (11). The results show similar distributions, confirming the validity of our method. **B)** The distribution of copy number states in patients (red) compared to that of controls (green), showed no significant differences.

In addition, we experimentally validated copy number states inferred from ZRPKM for exons 10-24 by relative quantification of the STR-PSV located at intron 12. We selected 11 individuals carrying predicted deletions and five individuals with predicted duplications. The global correlation between both methods was 80%, being perfect for duplications, but lower (72%) for deletions. This discordance could be due to the presence of additional partial copies containing intron 12 or incorrect relative quantification of the STR-PSV due to an amplification bias.

Definition of PSVs in *CNTNAP3*

In order to distinguish rearrangements affecting complete copies from polymorphic variation of partial copies, we started by defining PSVs that could uniquely tag the two functional copies of *CNTNAP3*. Previous expression data from GTEx already showed that the two complete copies had different expression patterns, with *CNTNAP3B* being expressed in brain and *CNTNAP3A* in blood, as well as in brain. By analyzing sequencing data form these two tissues, we have specifically assign PSVs to each copy. We compared the expression in both tissues of 27 positions differing between the two copies, according to the human reference genome hg19 version. In the human fetal brain samples, 15/27 positions were polymorphic, with both variants being expressed. In contrast, in the PBMC samples, the 15 PSVs with polymorphic expression brain were monomorphic, with only the variant

belonging to *CNTNAP3A* being expressed. Our data corroborates that *CNTNAP3A* is expressed both in PBMCs and brain, whereas *CNTNAP3B* is restricted to brain and is also expressed during the fetal period (Supplementary Table 6).

PSV ratios of the 15 positions differing between *CNTNAP3A* and *CNTNAP3B* showed values consistent with their expected ratio, being specific of *CNTNAP3A* but common between *CNTNAP3B* and partial copies. This suggests that partial copies originated from *CNTNAP3B*. To identify unique PSVs that specifically tag non-expressed partial copies, we performed multiple alignment of all paralogous copies and select those positions that were neither expressed in blood or fetal brain samples. We identified 2 PSVs in exon 7, originally annotated as belonging to *CNTNAP3B*, that showed ratios consistent with specific tagging of the partial copy containing exons 2-9 (Supplementary Figure 3). We did not identify any specific PSV tagging *CNTNAP3B* and the second largest partial copy containing exons 10-29.

Rearrangements affecting functional copies

To identify rearrangements affecting functional copies, we used the combined information of read depth and PSV ratios to infer the total integer copy number state and the number of copies for each paralogous copy. This approach was only possible for exons 2-9, where PSVs uniquely tagging *CNTNAP3A* and the partial copy were identified. The number of copies of *CNTNAP3B* was deducted by subtracting the number of previously deduced copies from the total copy number. Since this approach is based on relative quantification, it is limited by the expected high degree of conversion between copies and the high rate of polymorphic variation of partial copies, which can confound the results (Supplementary Figure 4). Consequently, further analyses are required to distinguish between CNVs in functional copies and conversion events.

Since complete copies contain all exons (1-24), deletions and duplications affecting a functional copy should show a compatible copy number integer state in exons 2-9 and 10-24. Therefore, we analyzed the total number of copies in exons 10-24. Since the correlation of inferred copy number states form ZRPKM and experimental approaches was 80%, we also quantified the total number of copies by the previously described PSV-STR located at intron

12. The results showed two deletions showing compatible values in inferred copy number state and PSV-STR ratios in two patients (AUT319 and GF1378), but none in controls (Figure 7A). Segregation studies in a patient with available parental samples (Figure 7A) showed that the deletion was inherited from the unaffected father. Regarding gains, three duplications of *CNTNAP3B* were found in two patients and a control (Table 4).

To ascertain the frequency of rearrangements in cases and controls we screened an additional cohort of patients (n=133) and controls (n=178) by the relative quantification of the PSV-STR located at intron 12. In addition, we designed a custom MLPA panel that enables the absolute quantification of copy number to help interpret doubtful cases by PSV-STR due to the limitations of relative quantification. The custom panel was first tested in samples with previously identified rearrangements using samples with the most frequent copy number state in both intervals (a total of six copies) as control references. Using that approach, we validated all previously identified rearrangements, with a mean value for deletions of 0.74 and 1.16 for duplications (Supplementary Table 7).

In the extended cohort of cases and controls, we detected 4/133 patients with deletions and 2/133 with duplications of *CNTNAP3B*. Regarding controls, 3/179 and 2/179 of individuals carried a deletion and duplication, respectively (Table 4).

	Targeted sequencing		***MLPA/PSV-STR***		**Total**	
	Cases	Controls	Cases	Controls	Cases	Controls
Del	2/216	0/64	4/133	3/179	6/349 (1.7%)	3/243 (1.2%)
Dup	2/216	1/64	2/133	2/179	4/349 (1.1%)	3/243(1.2%)
All	4/216	1/64	6/133	3/179	10/349 (2.9%)	6/243 (2.4%)

Table 4. Frequency of gains and losses of *CNTNAP3B* among cases and controls detected by two methods: targeted sequencing and experimental validation by MLPA and PSV-STR located at exon 11.

Segregation studies were performed in all patients with available parental samples (n=4) by PSV-STR quantification and, interestingly, we identified a *de novo* deletion in a patient (AUT186) (Figure 7B). The rest of rearrangements (n=3) were inherited from unaffected progenitors. No

parental samples of control individuals were available to perform segregation studies.

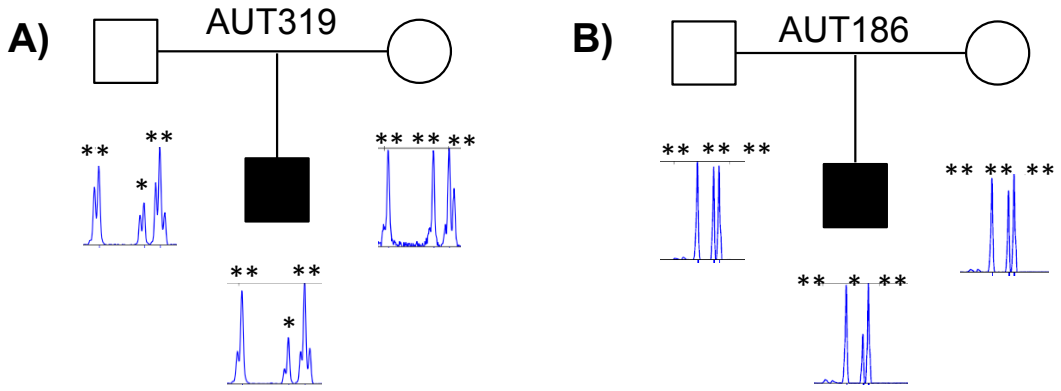


Figure 7. A) AUT319 presented a deletion of *CNTNAP3B*, identified by targeted sequencing, which was inherited from his unaffected father. **B)** *De novo* deletion of *CNTNAP3B* detected by PSV-STR and the MLPA custom panel.

Taking together results from both methods, the frequency of deletions did not differ significantly between patients and controls (p=0.74, Fisher's exact test) (Table 4). The frequency of duplications was also equal between patients (4/341, 1.2%) and controls (2/225, 1.2%) (p=1, Fisher's exact test).

Rare SNVs affecting coding regions of *CNTNAP3A* and *CNTNAP3B*

Finally, we studied SNVs in the coding regions of *CNTNAP3A* and *CNTNAP3B*. As previously done for the 7q11.23 region, we excluded SNVs present in more than one individual and the nature of the changes was assessed with respect to *CNTNAP3A*. In patients, we identified 35 unique variants, present in 32 individuals, compared to 14 variants in controls present in 12 individuals (Supplementary Table 9). All SNVs presented AB ratios consistent with a heterozygous state in multi-copy regions (mean AB ratio=12,5%, range=7-21%), with the range probably due to the varying number of underlying copies of each exon.

To determine if the identified rare SNVs were present in *CNTNAP3A*, which is expressed in blood, we performed RT-PCR followed by Sanger Sequencing of three patients with available RNA (Supplementary Table 8). Only one of the five analyzed variants was detected in blood mRNA, a

missense variant (p.C969Y) predicted as damaging affecting a highly conserved cysteine residue which forms a disulfide bridge with another nearby cysteine. In the rest of individuals with no available RNA, we validated and performed segregation studies of all LoF and missense damaging mutations. All were inherited from non-affected progenitors.

Interestingly, one of the five patients sequenced (AUT190), presented an abnormal pattern of expression, showing expression of PSVs belonging to both functional copies (*CNTNAP3A* and *CNTNAP3B)*. This pattern was not observed in 40 samples from PBMCs, in which the expression of *CNTNAP3B* was restricted to brain (Figure 8).

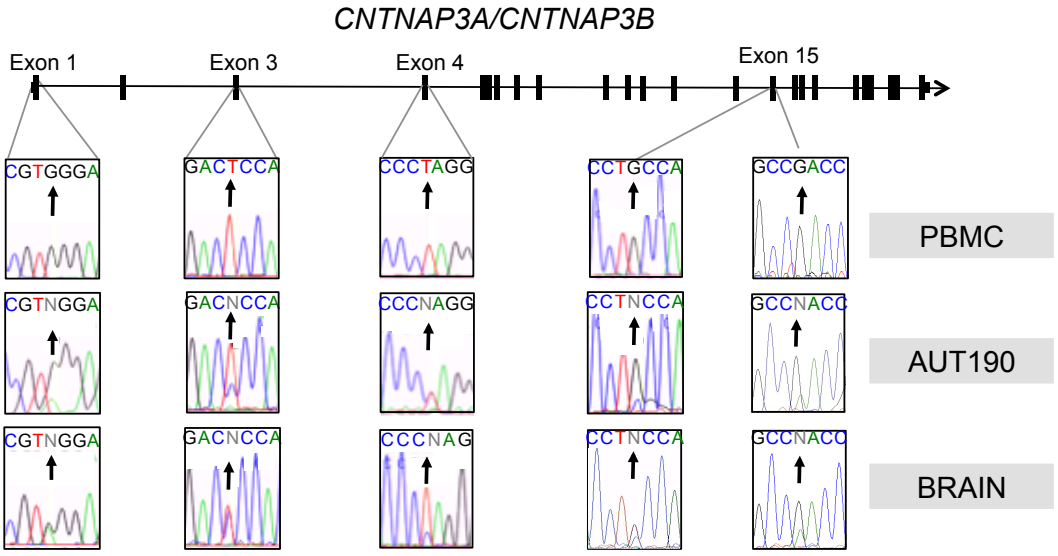


Figure 8. Expression of PSVs in brain, PBMCs from a control and PBMCs from patient AUT190.01. Brain samples PSVs from both copies of CNTNAP3 (*CNTNAP3A* and *CNTNAP3B*), whereas PBMC from 47 controls expressed only PSVs belonging to *CNTNAP3A*. In contrast, AUT190.01 expressed PSVs from both copies, as his unaffected mother.

To see if this could be due to gene conversion between the two functional copies, we analyzed PSVs ratios, but detected no evidence of gene conversion. Another possible explanation for this abnormal pattern of expression may be a structural rearrangement causing a positional effect that activates the expression of *CNTNAP3B* in blood. Sanger sequencing from blood mRNA of parents showed that the mother presented the same aberrant pattern of expression. Interestingly, this individual was found to be a carrier of

the common pericentromeric inversion of chromosome 9, inherited from his mother. However, the same pattern of expression was not detected in other carriers of the inversion.

Discussion

In this study we have analyzed the contribution to ASD of copy number and sequence variation in two complex regions of the genome: the 7q11.23 locus and the pericentromeric region of chromosome 9. Both regions have been specifically duplicated in the human lineage and contain brain-expressed genes, suggesting a role in the emergence of human traits. Therefore, their dysregulation could contribute to explain part of the missing heritability of ASD. Despite their possible role in neurodevelopmental disorders, they have been poorly studied due to its complexity. Here, we have developed a method to identify paralog-specific CNVs from targeted sequencing data, by analyzing distribution of multiple aligned reads across samples and relative PSV quantification. Using a sequencing approach, also allowed us to look for point mutations in coding regions, by adapting calling parameters to expected ratios in multi-copy regions. Finally, we have experimentally validated all results and performed additional studies in an extended cohort of patients and controls.

In the 7q11.23 locus, we have sequenced the two main candidate genes for the neurocognitive profile observed in WBS and its reciprocal duplication syndrome: *GTF2I* and *GTF2IRD2*. Our sequencing approach enabled us to evaluate both CNV and SNV variation. By the combined analysis of depth of coverage and experimental replication in an extended cohort, we detected an increased frequency of B block duplications, but not deletions, in ASD patients (2%) compared to controls (1%), but the difference was not statistically significant. However the detailed molecular characterization of the CNVs revealed a significant difference in the type of CNVs between both cohorts, with an increased frequency of duplications encompassing only block B in cases compared to controls ($p=0.02$). Therefore, our results point towards the contribution of duplications of block B duplications to the susceptibility of ASD.

To explore the mechanistic effects of these rearrangements, we characterized its origin and functional gene content. Analysis of PSVs

suggested that the rearrangement was compatible with a partial duplication of a telomeric or medial B block, with an unknown effect on functional gene content. The resulting rearrangement could either contain a partial non-functional copy with no effect or interrupt a functional copy of GTF2IRD2, given that the breakpoint seems to be located between exon 5 and 3 of this gene. In addition, block B duplications could also exert a positional effect by altering the expression of nearby genes. The only functional gene contained in the medial and telomeric block is *GTF2IRD2. GTF2IRD2* has been described a modulator of the WBS phenotype. Individuals with WBS carrying the larger 1.8 Mb deletions show higher cognitive impairments, obsessions and maladaptive behaviors than those with the most frequent 1.5 Mb deletion (22). This suggests that *GTF2IRD2* may be the locus behind this effect. However, further detailed characterization of the deletion breakpoint is needed to confirm this hypothesis.

In addition to the specific quantification of paralogous copies, our sequencing approach allowed us to study the role of point mutations. To this end, we adapted variant calling to detect SNVs in multi-copy regions with lower AB ratios. Since complex regions have been poorly studied and variation on such regions is not usually reported, we focused on novel unique variants. In *GTF2I*, no differences in the frequency of rare variants were observed between cases and controls. To show that variant calling on multiple mapping reads can successfully identify SNVs in duplicated regions, we validated and perform segregation studies on a set of selected deleterious variants, showing that all were inherited from non-affected progenitors. Due to the high degree of identity between copies and the lack of PSVs within read length, it was difficult to assign variants to their specific paralogous copies. For the unique LoF variant found in a patient affecting *GTF2I*, we performed a long-range PCR followed by Sanger sequencing, showing that it was not present in the only functional copy of the gene. Interestingly, the same filtering analysis applied to the coding regions of *GTF2IRD2* revealed a single rare variant in a patient (p.M1?) affecting the translation initiation site. Segregation studies showed that it is present in the proband, his affected sibling and inherited from his unaffected father. The mutation is predicted to alter the protein structure by abolishing the translation initiation site and switching it to an in-frame alternative start codon at a nearby methionine residue (Met21). Further studies assessing the translation and protein levels of the mutated

allele are needed to confirm its pathogenic effect. However, this pathogenic variant affecting *GTF2IRD2* adds to previous evidence showing an increased frequency of B block rearrangements in patients that pointed towards a role of *GTF2IRD2* in the susceptibility to ASD. Interestingly, all variants identified in this study are inherited from unaffected progenitors, suggesting that they may act as part of a multiple hit model, where several genetic variants may increase the risk but none is sufficient to cause it.

In the pericentromeric region of chromosome 9, we have sequenced all coding regions of *CNTNAP3,* a candidate gene for neurodevelopmental disorders, given its function and the fact that is specifically duplicated in humans. The region shows a highly complex genetic architecture that has changed dramatically over the latest human reference genomes. Therefore, its correct assembly and variation have not been established yet, making its study highly difficult. Our work provides the first basis for in-depth studies of the region. Our approach based on ZRPKM values replicated previous results, that showed a high variability in copy number (11). In addition, by analyzing windows overlapping with partial copies, we have shown that most variation of the region is due to rearrangements affecting non-functional partial copies. Moreover, we have shown that both functional copies of *CNTNAP3* are expressed in brain during the fetal period, a crucial time for the formation of the central nervous system, which is disrupted in ASD and other neurodevelopmental disorders (37–39). A part from brain, *CNTNAP3A* is also expressed in blood, providing an accessible tissue to explore the effect of variation in a functional copy. The expression screening of *CNTNAP3A* and *CNTNAP3B* has also allowed the functional validation of copy-specific PSVs, obtained by a detailed study of the underlying genetic sequence. Finally, the exploratory study by targeted sequencing has allowed the design of experimental approaches, setting the framework for further studies.

By the combined use of targeted sequencing data and experimental methods, we have validated and extended the results in an additional cohort of cases and controls. The results show that rearrangements affecting complete functional copies are relatively rare, with duplications being present in approximately 1% of cases and controls, while the frequency of deletions seemed slightly increased in patients compared to controls (1.7% vs 1.1%).

Interestingly, segregation studies in cases with available parental samples showed that one of the deletions had arisen *de novo,* suggesting a possible contribution to ASD susceptibility. However, further studies assessing the contribution of *de novo* variation in this region in ASD patients and controls are needed, given that duplicated regions have higher mutation rates than singly-copy regions, due to their highly repetitive content (40). In addition to the slightly increased frequency deletions in cases, we report a patient showing an aberrant pattern of expression, with expression of both functional copies (*CNTNAP3A* and *CNTNAP3B*) in blood. This patient carried the common pericentromeric inversion of chromosome 9, which was inherited from his mother, who presented the same aberrant pattern of expression in blood. This suggests that this effect may be mediated by a positional effect, activating the expression of *CNTNAP3B.* However, expression of *CNTNAP3B* was not observed in other individuals carrying the same inversion, implying that this patient may be carrying a cryptic rearrangement not detected with our current methods. Moreover, this patient was also found to carry three paternally inherited mutations, predicted as damaging. Since the variants were not detected in blood, they can either be present in the remaining functional copy (CNTNAP3B) or in one of the partial copies. These findings suggest a possible inherited second-hit in this patient. Therefore, taken together the results in this region warrant further studies to elucidate the possible role of *CNTNAP3* in neurodevelopmental disorders.

Therefore, our results show that sequencing-based studies can successfully detect changes in copy number of multi-copy regions and that they have similar detection rates compared to experimental methods. Moreover, they allow the study of multiple PSVs in a single assay, avoiding time-consuming protocols that analyze single positions at a time. Although most studies evaluating variation in complex regions have been carried out in WGS data, targeted sequencing offers some benefits over WGS approaches, as well as a few disadvantages. One of the main advantages is its higher coverage, which allows better discrimination of copy number states and PSV ratios. However, enrichment and pooling methods also introduce biases in coverage distribution between samples and targets, especially in GC rich regions. In this assay we included only coding sequences of candidate target genes and chose not to analyze PSVs lying in intronic sequences. Although this design was adequate for our purpose, capture of the entire sequence may be a

better approach for regions with less coding content or higher sequence identity. Finally, focusing on specific regions allows applying customized computational pipelines and specific experimental validation methods.

To conclude, our study shows that NGS based approaches can successfully identify copy number and sequence variants in complex regions, by providing information about the number of copies (by read-depth) and the identity of duplicated sequences (by PSVs). The development of new methodologies, such as the one described here, together with the refinement of the human reference genome, will increase the study of complex regions and elucidate their role in complex disorders. Here, we explored the variation and the role of two duplicated regions in ASD: the 7q11.23 region and the pericentromeric region of chromosome 9. Our results suggest that these two regions may be involved in ASD susceptibility, and warrant further detailed studies and molecular characterization in additional cases.

References

1. Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev. 2012 Jun;22(3):229–37.

2. Huddleston J, Eichler EE. An Incomplete Understanding of Human Genetic Variation. Genetics. 2016 Apr;202(4):1251–4.

3. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet. 2006 Jul;7(7):552–64.

4. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446–50.

5. Emanuel BS, Shaikh TH. Segmental duplications: an “expanding” role in genomic instability and disease. Nat Rev Genet. 2001;2(10):791–800.

6. Samonte RV, Eichler EE. Segmental duplications and the evolution of the primate genome. Nat Rev Genet. 2002;3(1):65–72.

7. Bailey J a, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet. 2006;7(7):552–64.

8. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet. 2005 Jul;77(1):78–88.

9. Goidts V, Cooper DN, Armengol L, Schempp W, Conroy J, Estivill X, et al.

Complex patterns of copy number variation at sites of segmental duplications: An important category of structural variation in the human genome. Hum Genet. 2006;120(2):270–84.

10. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, et al. Diversity of human copy number variation and multicopy genes. Science. 2010 Oct 29;330(6004):641–6.

11. Handsaker RE, Van Doren V, Berman JR, Genovese G, Kashin S, Boettger LM, et al. Large multiallelic copy number variations in humans. Nat Genet. 2015;(January):1–10.

12. Nuttle X, Itsara A, Shendure J, Eichler EE. Resolving genomic disorder-associated breakpoints within segmental DNA duplications using massively parallel sequencing. Nat Protoc. 2014;9(6):1496–513.

13. Itsara A, Vissers LELM, Steinberg KM, Meyer KJ, Zody MC, Koolen DA, et al. Resolving the breakpoints of the 17q21.31 microdeletion syndrome with next-generation sequencing. Am J Hum Genet. 2012 Apr 6;90(4):599–613.

14. Forni D, Martin D, Abujaber R, Sharp AJ, Sironi M, Hollox EJ. Determining multiallelic complex copy number and sequence variation from high coverage exome sequencing data. BMC Genomics. 2015;16(1):891.

15. Steinberg KM, Antonacci F, Sudmant PH, Kidd JM, Campbell CD, Vives L, et al. Structural diversity and African origin of the 17q21.31 inversion polymorphism. Nat Genet. 2012;44(8):872–80.

16. Abu Bakar S, Hollox EJ, Armour J a L. Allelic recombination between distinct genomic locations generates copy number diversity in human beta-defensins. Proc Natl Acad Sci U S A. 2009;106(3):853–8.

17. Cuscó I, Corominas R, Bayés M, Flores R, Rivera-Brugués N, Campuzano V, et al. Copy number variation at the 7q11.23 segmental duplications is a susceptibility factor for the Williams-Beuren syndrome deletion. Genome Res. 2008 May;18(5):683–94.

18. Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, et al. A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nat Genet. 2001 Nov;29(3):321–5.

19. Antonell A, Del Campo M, Magano LF, Kaufmann L, de la Iglesia JM, Gallastegui F, et al. Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. J Med Genet. 2010 May;47(5):312–20.

20. Morris C a, Mervis CB, Hobart HH, Gregg RG, Bertrand J, Ensing GJ, et al. GTF2I hemizygosity implicated in mental retardation in Williams syndrome: genotype-phenotype analysis of five families with deletions in the Williams syndrome region. Am J Med Genet A. 2003;123A(1):45–59.

21. Sakurai T, Dorr NP, Takahashi N, McInnes LA, Elder GA, Buxbaum JD. Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions. Autism Res. 2011 Feb;4(1):28–39.

22. Porter MA, Dobson-Stone C, Kwok JBJ, Schofield PR, Beckett W, Tassabehji M. A role for transcription factor GTF2IRD2 in executive function in Williams-Beuren syndrome. PLoS One. 2012;7(10):e47457.

23. Pérez Jurado LA, Wang YK, Peoples R, Coloma A, Cruces J, Francke U. A duplicated gene in the breakpoint regions of the 7q11.23 Williams-Beuren syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. Hum Mol Genet. 1998 Mar;7(3):325–34.

24. Tuzun E, Sharp AJ, Bailey J a, Kaul R, Morrison VA, Pertz LM, et al. Fine-scale structural variation of the human genome. Nat Genet. 2005;37(7):727–32.

25. Antonell A, de Luis O, Domingo-Roura X, Pérez-Jurado LA. Evolutionary mechanisms shaping the genomic structure of the Williams-Beuren syndrome chromosomal region at human 7q11.23. Genome Res. 2005 Sep;15(9):1179–88.

26. Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, et al. Gene copy number variation spanning 60 million years of human and primate evolution. Genome Res. 2007;17(9):1266–77.

27. Fortna A, Kim Y, MacLaren E, Marshall K, Hahn G, Meltesen L, et al. Lineage-specific gene duplication and loss in human and great ape evolution. PLoS Biol. 2004;2(7).

28. Wilson GM, Flibotte S, Missirlis PI, Marra MA, Jones S, Thornton K, et al. Identification by full-coverage array CGH of human DNA copy number increases relative to chimpanzee and gorilla. Genome Res. 2006;16(2):173–81.

29. Südhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. Nature. 2008 Oct 16;455(7215):903–11.

30. Spiegel I. Caspr3 and Caspr4, Two Novel Members of the Caspr Family Are Expressed in the Nervous System and Interact with PDZ Domains. Mol Cell Neurosci. 2002 Jun;20(2):283–97.

31. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. Science. 2015 May 8;348(6235):660–5.

32. Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome deletions. Am J Hum Genet. 2003 Jul;73(1):131–51.

33. Del Campo M, Antonell A, Magano LF, Muñoz FJ, Flores R, Bayés M, et al. Hemizygosity at the NCF1 gene in patients with Williams-Beuren syndrome decreases their risk of hypertension. Am J Hum Genet. 2006;78(4):533–42.

34. Casimir CM, Bu-Ghanim HN, Rodaway AR, Bentley DL, Rowe P, Segal AW. Autosomal recessive chronic granulomatous disease caused by deletion at a dinucleotide repeat. Proc Natl Acad Sci U S A. 1991 Apr 1;88(7):2753–7.

35. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of

image analysis. Nat Methods. 2012 Jul;9(7):671–5.

36. Cantsilieris S, Western PS, Baird PN, White SJ. Technical considerations for genotyping multi-allelic copy number variation (CNV), in regions of segmental duplication. BMC Genomics. 2014;15:329.

37. Stoner R, Chow ML, Boyle MP, Sunkin SM, Mouton PR, Roy S, et al. Patches of disorganization in the neocortex of children with autism. N Engl J Med. 2014 Mar 27;370(13):1209–19.

38. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013;155(5):1008–21.

39. Willsey a. J, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle R a., et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013;155(5):997–1007.

40. Campbell CD, Eichler EE. Properties and rates of germline mutations in humans. Trends Genet. 2013 Oct;29(10):575–84.

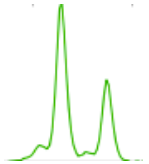
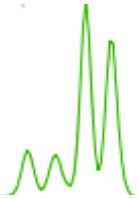
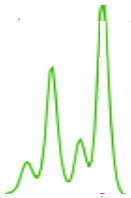
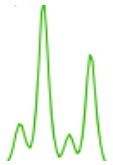
Supplementary Information

Supplementary Table 1. Sequences of primers and MLPA probes used for validation of results regarding *CNTNAP3*.

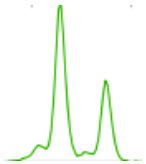
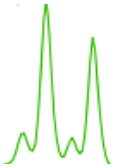
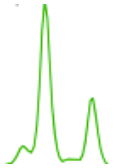
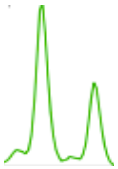
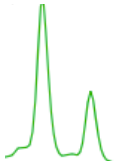
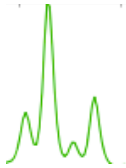
PSV-STR	
Expected size (hg19)	**Sequence**
Partial copy (65 Mb): 328 bp CNTNAP3A: 342bp CNTNAP3B: 338 bp	**Forward primer**
	5'-[6FAM]GGTTCTCACAGTGAATGTTGGAGAAAAATCTCCT-3'
	Reverse primer
	5'-TGTATTAGGTCTGAGTCCTTTAGGGAGCC-3'
MLPA	
CNTNAP3A	**L-probe:** GAATTCCGACACCCGGAAGATCGCGGGCATTCA
	R-probe: GACGTAAAAGCTTTCATGGGTGTTTAGAA
CNTNAP3B and partial copy containing exons 2-9	**L-probe:** GAATTCCGACACCCGGAAGATCGCGGGCATTCA
	R-probe: CACGTAAAAGCTTTCATGGGTGTTTAGAA
CNTNAP3A, CNTNAP3B and partial copy located at Mb 65	**L-probe:** GGCTTTACTATATTGATGCAGATGG
	R-probe: AAGTGGCCCCCTGGGACCATTTCTTGTGTACTGCAAT

ID	Status	CNV-type	# of blocks (C\A\B)	Block C	Block A	B block		Block origin	
				BCSTR1 (Cc:Cm:Ct)	BASTR1 (WS2)	BBSTR1	NCF1 (PG/G)	SSN11a	PSV ratios
AUT1_01	Patient	DUP	NA \ 7 \ 7	NA	138, 138, 149, 151, 151, 151, 154,	130, 130, 133, 133, 133, 135, 152	5/2	2:3	NA
AUT61_01	Patient	DUP	6 \ 6 \ 7	3:1:2	138, 138, 142, 149, 151, 153	130, 134, 134, 136, 140, 144, 153	5/2	2:2	NA
AUT111_01	Patient	DUP	6 \ 7 \ 7	3:2:1	134, 138, 151, 151, 151, 151, 153	130, 134, 134, 136, 136, 142, 148	5/2	2:2	NA
AUT154_01	Patient	DUP	6 \ 6 \ 7	2:2:2	135, 139, 139, 151, 151, 153,	130, 134, 134, 134, 140, 140, 153	5/2	2:2	NA
AUT192_01	Patient	DUP	6 \ 6 \ 7	2:2:2	141, 141, 145, 154, 154, 158	131, 135, 135, 135, 139, 139, 141	5/2	1:3	Medial/telomeric
GF2176	Patient	DUP	6 \ 6 \ 7	2:3:1	137, 141, 145, 154, 154, 156	135, 135, 135, 135, 137, 153, 155	5/2	1:3	Medial/telomeric
GF0956	Patient	DUP	6 \ 6 \ 7	3:2:1	137, 141, 145, 154, 152, 154	131, 135, 135, 135, 137, 141, 143	4/3	2:2	Medial/telomeric
GF1386	Patient	DUP	6 \ 6 \ 7	2:2:2	137, 141, 154, 154, 156, 156	131, 131, 135, 135, 137, 141, 154	5/2	2:2	Medial/telomeric
GF1218	Patient	DEL	5 \ 5 \ 5	2:3:0	137, 137, 154, 156, 158	135, 137, 139, 154, 157	4/3	2:2	Centromeric
GF1349	Patient	DEL	5 \ 5 \ 5	2:1:2	137, 141, 145, 154, 156	131, 135, 137, 146, 154	2/3	2:2	Centromeric
AUT193.01	Patient	DEL	5 \ 5 \ 5	2:2:2	141, 141, 154, 154, 158	131, 131, 135, 143, 154	3/2	2:2	NA
AUT26_01	Patient	DEL	6 \ 6 \ 5	2:2:2	139, 139, 143, 150, 151, 154	134, 134, 136, 136, 155	3/2	1:2	NA
AR123	Control	DUP	7 \ 7 \ 7	2:3:2	144, 144, 160, 158, 158, 156, 148	134, 136, 138, 138, 140, 146, 158	5/2	1:3	NA
CE8998241	Control	DUP	7 \ 7 \ 7	2:2:3	136, 140, 154, 152, 152, 148, 144	130, 136, 136, 136, 138, 144, 152	5/2	2:3	NA
CE9683608	Control	DUP	7 \ 7 \ 7	1:2:4	140, 140, 154, 152, 152, 152, 148	130, 130, 134, 134, 134, 142, 158	5/2	2:2	NA
CAT52	Control	DUP	7 \ 7 \ 7	3:2:2	144, 144, 148, 156, 158, 158, 160	134, 136, 138, 138, 140, 144, 158	5/2	2:2	NA
CAT56	Control	DUP	6 \ 7 \ 7	2:2:2	144, 144, 158, 158, 156, 148, 144	134, 134, 138, 138, 140, 150, 160	5/2	2:2	NA
SW162.11	Control	DUP	6 \ 6 \ 7	2:2:2	136, 140 154, 154, 152, 144	130, 134, 134, 138, 142, 152, 156	4/3	2:2	NA
SW48.12	Control	DUP	7 \ 7 \ 7	2:3:2	140, 140, 154, 154, 152, 152, 152	128, 130, 134, 134, 134, 136, 150	5/2	2:2	NA
SW223.12	Control	DUP	6 \ 6 \ 7	2:2:2	139, 139, 152, 152, 154, 154	NA		2:2	NA
SW578.11	Control	DUP	6 \ 7 \ 7	2:2:2	141, 141, 141, 145, 153, 153, 156	131, 131, 135, 135, 135, 139, 415	5/2	2:2	NA
SW643.11	Control	DUP	6 \ 6 \ 7	3:2:1	136, 141 144, 144, 151, 153	135, 135, 135 ,135, 135 ,146, 150	5/2	1:3	NA
SW152-11	Control	DEL	5 \ 5 \ 5	2:1:2	140, 140, 154, 152, 152	130, 130, 134, 142, 150	3/2	2:2	NA
SW271-12	Control	DEL	6 \ 5 \ 5	2:2:2	136, 136, 154, 144, 144	132, 134, 136, 138, 140	2/3	2:2	NA
SW62.12	Control	DEL	6 \ 5 \ 5	2:2:2	140, 144, 147, 147, 155	134, 138, 142, 144, 148	3/2	2:2	NA
SW74.12	Control	DEL	5 \ 5 \ 5	1:2:2	140, 144, 158, 156, 154	138, 140, 144, 148, 156	2/3	1:3	NA
AR100	Control	DEL	5 \ 5 \ 5	1:2:2	140, 144, 158, 156, 154	138, 138, 140, 142, 154	3/2	NA	NA
CE6898617	Control	DEL	6 \ 5 \ 5	2:2:2	136, 140, 152, 152, 144	130, 134, 134, 150, 154	3/2	NA	NA
SW12.02	Control	DEL	5 \ 5 \ 5	1:2:2	140, 140, 152, 154, 156	134, 134, 138, 148, 156	2/3	1:3	NA
SW558.12	Control	DEL	5 \ 5 \ 5	2:2:1	136, 140, 153, 155, 155	129, 131, 135, 137, 141	3/2	2:2	NA
SW631.12	Control	DEL	5 \ 5 \ 5	2:2:1	136, 140, 153, 153, 155	131, 135, 137, 201, 158	3/2	NA	NA
SS2_11	Control	DEL	5 \ 5 \ 5	2:1:2	139, 139, 146, 146, 154	133, 135, 135, 141, 141	1/4	2:2	Centromeric

Supplementary Table 2. Characterization of all identified rearrangements

	PSV		Expected ratio			Individual		
						SS2.11	GF1349	GF1218
NCF1	**2 bp deletion in Bc and Bt blocks**		Control ratio  4/2 Ratio Bc+Bt/Bm=1.95			 1/4 Ratio=0.26	 2/3 Ratio=0.69	 3/2 Ratio=1.47
GTF2IRD2	**Position**	**Exon**	**Centromeric**	**Medial**	**Telomeric**	**SS2_11**	**GF1349**	**GF1218**
	7_74211003	**e16**	0,80	0,60	0,60	0,74	0,79	0,85
	7_74211416	**e16**	0,40	0,20	0,40	0,34	0,36	0,40
	7_74211558	**e16**	0,80	0,60	0,60	0,83	0,81	0,78
	7_74211784	**e16**	0,40	0,20	0,40	0,35	0,42	0,36
	7_74211879	**e16**	0,80	0,60	0,60	0,77	0,80	0,81
	7_74212016	**e16**	0,60	0,60	0,80	0,53	0,54	0,60
	7_74212026	**e16**	0,60	0,60	0,80	0,54	0,55	0,62
	7_74212186	**e16**	0,80	0,60	0,60	0,83	0,81	0,81
	7_74212418	**e16**	0,40	0,20	0,40	0,44	0,41	0,37
	7_74212544	**e16**	0,80	0,60	0,60	0,89	0,83	0,86
	7_74212546	**e16**	0,80	0,60	0,60	0,89	0,83	0,85
	7_74212596	**e16**	0,80	0,60	0,60	0,89	0,82	0,83
	7_74217714	**e14**	0,80	0,60	0,60	0,77	0,78	0,78
	7_74237306	**e5**	0,80	0,60	0,60	0,78	0,82	0,79

Supplementary Table 3. Deletions identified by targeted sequencing were validated by the *NCF1*-PSV and assigned to the centromeric copy by means of the PSV ratios of *GTF2IRD2*.

	PSV	Expected ratio			Individual				
					GF0956	GF1386	GF2176	AUT192.01	C2
NCF1	**2 bp deletion in Bc and Bt blocks**	Control ratio  4/2 Ratio=1.95			 4/3 1.28	 5/2 2.52	 5/2 2.43	 5/2 2.30	 5/2 2.38
GTF2IRD2	**Position**	**Exon**	**Centromeric**	**Medial**	**Telomeric**	**GF0956**	**GF1386**	**GF2176**	**AUT192.01**	**C2**
	7_74211003	**e16**	0,57	0,71	0,71	0,73	0,61	0,66	0,63	0,61
	7_74211416	**e16**	0,29	0,43	0,29	0,43	0,32	0,32	0,32	0,23
	7_74211558	**e16**	0,57	0,71	0,71	0,68	0,68	0,67	0,70	0,56
	7_74211784	**e16**	0,29	0,43	0,29	0,27	0,34	0,35	0,36	0,28
	7_74211879	**e16**	0,57	0,71	0,71	0,86	0,84	0,71	0,67	0,58
	7_74212016	**e16**	0,71	0,71	0,57	0,71	0,61	0,66	0,62	0,72
	7_74212026	**e16**	0,71	0,71	0,57	0,70	0,60	0,68	0,63	0,74
	7_74212186	**e16**	0,57	0,71	0,71	0,73	0,64	0,73	0,64	0,61
	7_74212418	**e16**	0,29	0,43	0,29	0,36	0,37	0,38	0,32	0,24
	7_74212544	**e16**	0,57	0,71	0,71	0,71	0,69	0,69	0,64	0,64
	7_74212546	**e16**	0,57	0,71	0,71	0,71	0,70	0,69	0,64	0,63
	7_74212596	**e16**	0,57	0,71	0,71	0,71	0,74	0,70	0,67	0,71
	7_74217714	**e14**	0,57	0,71	0,71	0,67	0,59	0,63	0,68	0,53
	7_74237306	**e5**	0,57	0,71	0,71	0,69	0,68	0,65	0,68	0,63

Supplementary Table 4. Duplications identified by targeted sequencing were validated by the NCF1-PSV, but could not be assigned to a specific copy by means of the PSV ratios of GTF2IRD2. An individual with a known centromeric

Individual	Status	Chr	Start	Ref	Alt	AAChange.refGene	AB ratio	Position of nearby PSV	Imputed copy	Inheritance
GF1149	AUT	7	74143172	C	T	GTF2I:NM_001518:p.P290L	18%			NA
AUT294-01	AUT	7	74143199	A	G	GTF2I:NM_001518:p.N299S	12%			NA
AUT178-01	AUT	7	74143234	G	A	GTF2I:NM_001518::p.E311K	9%	74143252	Telomeric	Maternal
AUT294-01	AUT	7	74146856	C	A	GTF2I:NM_001518:p.P345Q	18%	74146870	NA (not within read length)	NA
GF1494	AUT	7	74146939	0	A	GTF2I:NM_001518:p.Y373_G374delinsX	16%		Centromeric or telomeric	NA
SW271-12	control	7	74146952	G	A	GTF2I:NM_001518:p.R377H	28%			NA
SS41-11	control	7	74150842	C	T	GTF2I:NM_001518:p.A438V	22%			NA
AUT157-01	AUT	7	74150947	G	A	GTF2I:NM_001518::p.R473Q	11%			Maternal
SW278-11	control	7	74150991	C	T	GTF2I:NM_001518:p.Q488X	25%			NA
AUT229-01	AUT	7	74157805	G	A	GTF2I:NM_001518:p.V525I	11%			NA
SS15-12	control	7	74160696	C	T	GTF2I:NM_001518:p.P631S	14%			NA
AUT221-01	AUT	7	74160733	T	C	GTF2I:NM_001518:p.I643T	18%			NA
GF1439	AUT	7	74162417	G	A	GTF2I:NM_001518:p.V671I	12%			NA
AUT225-01	AUT	7	74163629	C	G	GTF2I:NM_001518:p.A713G	9%			Maternal
SW141-11	control	7	74163640	C	G	GTF2I:NM_001518:p.P717A	16%			NA
SW203-11	control	7	74163709	C	T	GTF2I:NM_001518:p.R740X	14%			NA
AUT218-01	AUT	7	74168252	G	A	GTF2I:NM_001518:p.V865I	14%			Maternal

Supplementary Table 5. Rare SNVs identified in the coding regions of *GTF2I* in patients and controls. All mutations show AB ratios compatible with their presence in multi-copy regions.

CNTNAP3A_hg19	CNTNAP3A_hg19_base	CNTNAP3B_hg19	CNTNAP3B_hg19_base	Exon	Brain (n=3)	PBMC (n=47)
chr9:39287993	C	43685363	A	Exon 1	G/A	G
chr9:39239167	A	43737343	C	Exon 3	T/C	T
chr9:39193171	A	43800965	C	Exon 4	T/C	T
chr9:39178312	C	43815830	C	Exon 5	G/C	G
chr9:39133127	C	43800965	C	Exon 13	G/C	G
chr9:39109245	C	43884984	C	Exon 15	G/C	G
chr9:39109210	C	43885019	A	Exon 15	G/A	G
chr9:39103904	G	43890303	T	Exon 16	C/T	C
chr9:39103796	G	43890411	T	Exon 16	C/T	C
chr9:39103763	A	43890444	C	Exon 16	T/C	T
chr9:39102696	G	43891510	A	Exon 17	C/A	C
chr9:39088546	G	43905702	T	Exon 19	C/T	C
chr9:39088489	T	43905759	T	Exon 19	A/T	A
chr9:39086776	G	43907474	T	Exon 20	C/T	C
chr9:39086728	A	43907522	C	Exon 20	T/C	T

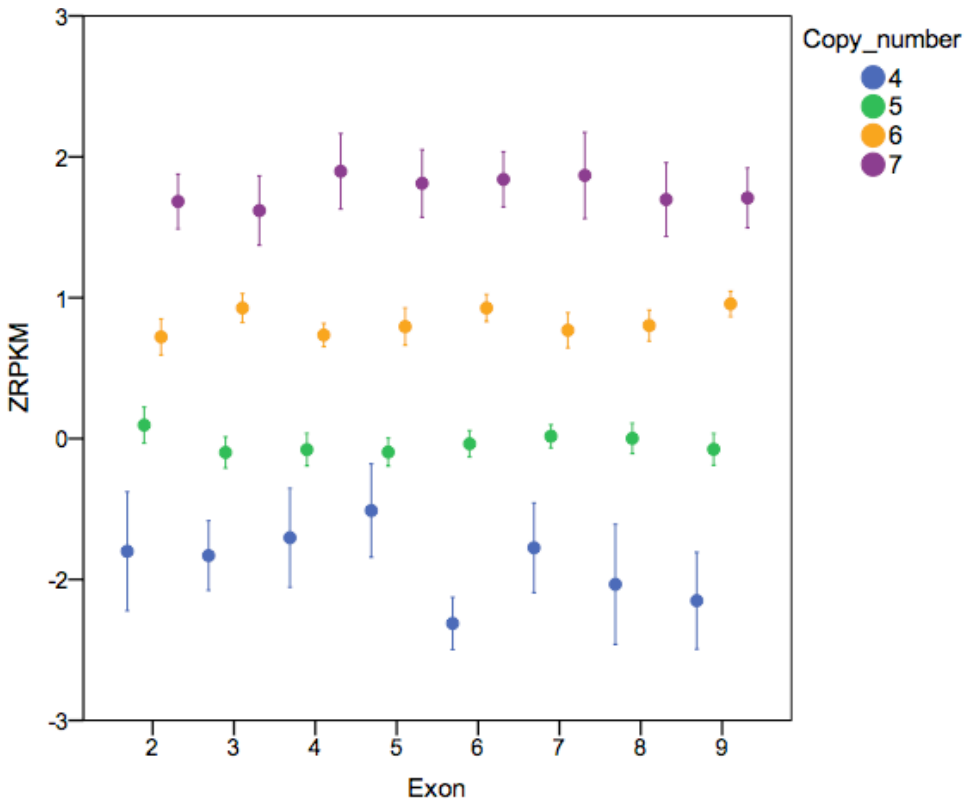
Supplementary Table 6. Expression of PSVs according to the hg19 version of the human genome in two PBMCs and brain

	Exons 2-9						Exons 10-24		
	MLPA		Targeted sequencing				MLPA	Targeted sequencing	
Individual	Exon7_CNTNAP3A	Exon7_CNTNAP3B_partial_copy40	Inferred copy number	#CNTNAP3B	#CNTNAP3A	#partial copy	Exon12_all_copies	Inferred copy number	Rearrangement
AUT319_01	0,96	0,59	4	1	2	1	0,91	5	*CNTNAP3B* deletion
GF1378	0,93	0,59	4	1	2	1	0,87	5	*CNTNAP3B* deletion
AUT158_01	1,08	0,97	7	3	2	2	1,23	7	*CNTNAP3B* duplication
GF0909	1,09	1,11	7	3	2	2	1,17	7	*CNTNAP3B* duplication
SS8_11	1,03	1,14	7	3	2	2	1,23	7	*CNTNAP3B* duplication

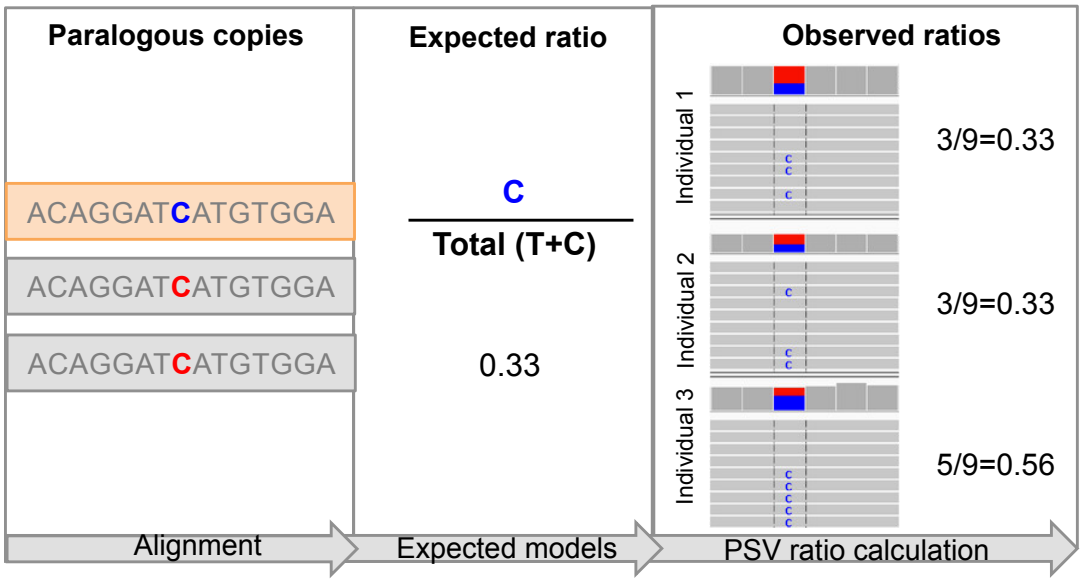
Supplementary Table 7. Validation of rearrangements by the custom panel MLPA

Individual	Variant	Gene	Protein effect	Nearby PSV	AB	Expression in blood	Segregation
AUT163.01	9:39099997_C_T	*CNTNAP3*	p.C969Y		0,11	Yes	NA
AUT190.01	9:39073898_-_T	*CNTNAP3*	p.E1286fs		0,11	No	Paternal
AUT190.01	9:39149837_C_G	*CNTNAP3*	p.D539H		0,12	No	Paternal
AUT190.01	9:39176082_A_G	*CNTNAP3*	p.F312S		0,13	No	Paternal
AUT208.01	9:39078719_C_G	*CNTNAP3*	p.R1214P		0,15	No	NA
AUT221.01	9:39088558_C_G	*CNTNAP3*	p.V1028L	39088546	0,12	NA	NA
AUT231.01	9:39073931_-_T	*CNTNAP3*	c.3822dupA		0,1	NA	Maternal
AUT269.01	9:39078881_G_A	*CNTNAP3*	p.A1160V		0,19	NA	NA
AUT294.01	9:39171380_C_T	*CNTNAP3*	p.R440K		0,12	NA	NA
AUT306.01	9:39078698_G_A	*CNTNAP3*	p.A1221V		0,15	NA	NA
GF0900	9:39239106_T_C	*CNTNAP3*	p.T92A		0,18	NA	NA
GF0922	9:39266922_G_A	*CNTNAP3*	p.P56L		0,09	NA	NA
GF0939	9:39086729_G_A	*CNTNAP3*	p.A1113V	39086728	0,19	NA	NA
GF0949	9:39176073_A_G	*CNTNAP3*	p.I315T		0,12	NA	NA
GF0963	9:39099988_C_T	*CNTNAP3*	p.G972E		0,21	NA	NA
GF1038	9:39078779_C_T	*CNTNAP3*	p.R1194H		0,16	NA	NA
GF1064	9:39149803_CTGT_-	*CNTNAP3*	c.1646_1649del		0,1	NA	NA
GF1082	9:39238994_A_G	*CNTNAP3*	p.I129T		0,2	NA	NA
GF1188	9:39086820_C_G	*CNTNAP3*	p.D1083H	39086776	0,08	NA	NA
GF1277	9:39177454_AGGGTG_-	*CNTNAP3*	p.261_263del		0,15	NA	NA
GF1315	9:39118177_T_G	*CNTNAP3*	p.Q720H		0,12	NA	NA
GF1444	9:39166059_C_T	*CNTNAP3*	p.D450N		0,16	NA	NA
GF1451	9:39088558_C_T	*CNTNAP3*	p.V1028I	39088546	0,11	NA	NA
GF1487	9:39099965_T_-	*CNTNAP3*	p.R980fs		0,11	NA	NA
GF1534	9:39140636_C_T	*CNTNAP3*	c.1757-1G>A		0,09	NA	NA
GF1536	9:39074002_A_G	*CNTNAP3*	p.I1251T		0,11	NA	NA
GF1558	9:39144277_G_T	*CNTNAP3*	p.D572E		0,12	NA	NA
GF1592	9:39166025_T_C	*CNTNAP3*	p.K461R		0,13	NA	NA
GF1650	9:39193146_C_T	*CNTNAP3*	p.E173K	39193171	0,18	NA	NA
GF1672	9:39178210_C_G	*CNTNAP3*	p.G229A		0,18	NA	NA
GF1820	9:39102658_G_A	*CNTNAP3*	p.T864M	39102696	0,16	NA	NA
GF1898	9:39118131_A_G	*CNTNAP3*	p.Y736H		0,1	NA	NA

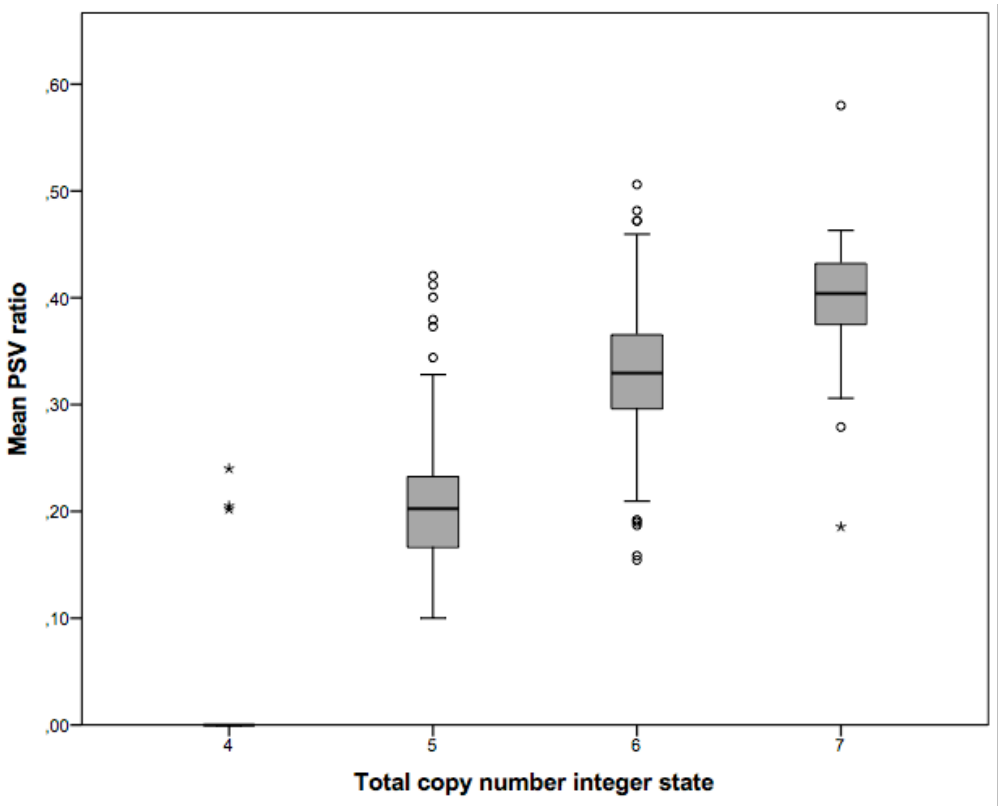
Supplementary Table 8. Rare SNVs identified in the coding region of *CNTNAP3*



Supplementary Figure 1. ZRPKM thresholds to assign copy number states in the first window (exons 2-9) are shown in the figure. Values were set to <-1.5 ZRPKM for a diploid copy number of 4, >-1.5 and <0 for a diploid copy number of 5, >0 and <1.5 for a diploid copy number of 6 and >1.5 for a diploid copy number of 7.

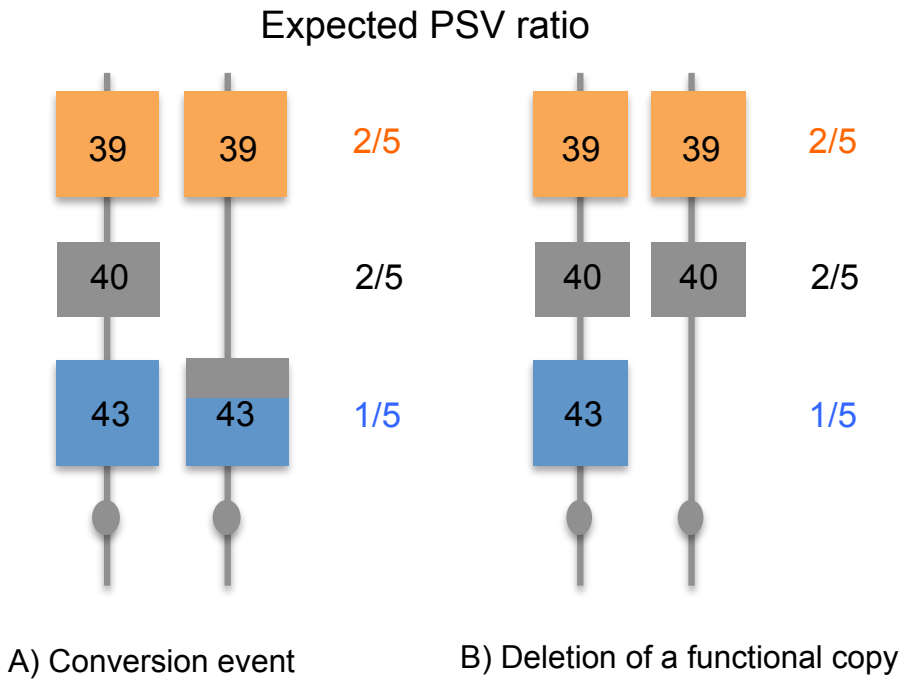


Supplementary Figure 2. Schematic representation of the method used to calculate PSV ratios.



	Total copy number integer state			
	4	**5**	**6**	**7**
Expected ratio	0/4 (0)	2/5 (0.2)	2/6 (0.33)	3/7 (0.43)
Mean	0.04	0.21	0.33	0.4

Supplementary Figure 3. Mean PSV ratio of the two PSVs in exon 7 specifically tagging the partial copy containing exons 2-9. The difference in mean PSV ratio between groups is statistically significant ($F_{3,25}$=153.790, p<0.0001). The table shows the expected PSV ratio assuming that variation in copy number is due to CNVs of the partial copy.



	CNTNAP3A PSV ratio	Partial copy 40 PSV ratio	Total copy number
A	0.2	0.2	5
B	0.2	0.2	5

Supplementary Figure 4. Schematic representation of two events affecting a partial (A) or a complete copy (B) indistinguishable by our approach. Figure A (left) represents a gene conversion between *CNTNAP3B* and the partial copy located at 40 Mb in an individual lacking a partial copy. The right panel B) depicts a deletion affecting *CNTNAP3B*. Both would result in the same PSV ratios and total copy number integer state, as shown in the table below, and would be indistinguishable by our approach.s

CHAPTER 4

Genetic modifiers of the Williams-Beuren syndrome neurobehavioral profile

Marta Codina-Solà, Debora Pérez-Garcia, Raquel Flores, Gabriela Palacios, Luis A. Pérez Jurado, Ivon Cuscó

In preparation

Abstract

Introduction: The hallmark of the neurobehavioral phenotype of Williams-Beuren syndrome (WBS) is increased sociability and relatively preserved language skills, often described as the opposite to Autism Spectrum Disorders (ASD). However, the prevalence of ASD among WBS individuals is 12%, approximately 10 fold higher than in the general population. To look for phenotypic modifying factors, we have performed an unbiased study of genetic variants in eight individuals with WBS and co-occurring ASD.

Subjects and Methods: The study was conducted on eight individuals with WBS. All had a confirmed diagnosis of ASD by the Autism Diagnostic Interview-Revised (ADI-R). We performed a detailed characterization of the deletion at the molecular level and looked for genome-wide variants, both rare and common, using exome sequencing.

Results: No evidence of second-hit *cis* or *trans* acting factors was found in the remaining hemizygous allele, neither of large Copy Number Variants. However, we identified several loss of function (LoF) mutations in ASD candidate genes, as well as in brain-expressed genes with a high probability of LoF intolerance. In addition, we found an increased number of rare deleterious variants in candidate genes in females compared to males.

Conclusions: Our results suggest that deleterious variants in functionally constrained brain expressed genes could act as second-hit modifying factors. The higher load of deleterious variants in ASD candidate genes in females suggests that the female protective effect may still exert a role the variable expressivity of disorders with full penetrance.

Introduction

Williams-Beuren syndrome (WBS, OMIM #194050) is a rare neurodevelopmental disorder resulting from a heterozygous deletion of 26-28 genes at chromosome 7q11.23, estimated to affect approximately 1 in 7500 individuals (1). The 7q11.23 chromosomal region has a complex genomic architecture that can predispose to Non Allelic Homologous Recombination (NAHR). The single copy region of the WBS locus is flanked by three large SDs, which share a high degree of homology (≈98-99%) located in the centromeric (c), medial (m) and telomeric (t) sides. There are two types of recurrent rearrangements that result in WBS syndrome: the most frequent is a 1.55Mb deletion that occurs in 87% of cases as a result of NAHR between the medial and centromeric B blocks. However, around 10% of patients present a larger deletion of 1.8Mb due to a crossing over between centromeric and medial A blocks. The remaining 3% of WBS patients show atypical deletions mediated by other mechanisms (2,3).

WBS is characterized by a set of clinical signs and symptoms, although the phenotype can vary greatly across patients. It is a multisystemic disorder characterized by cardiovascular disease with frequent supravalvar aortic stenosis and hypertension, distinctive facies, connective tissue abnormalities and growth and endocrine alterations among others. Regarding cognition and behavior, it is associated with mild to moderate intellectual disability (ID), an overfriendly personality and relative language preservation. Interestingly, the reciprocal duplication of 7q11.23 (OMIM #609757) results in a mirror phenotype, characterized by language impairment, speech delay and alterations and deficits in social interaction, which is often associated with Autism Spectrum Disorders (ASD) (4).

Due to their different associated behaviors, ASD and WBS have often been described as two diametric opposite disorders (5), although some authors have suggested that this consideration is an oversimplification of both phenotypes (6). In this regard, several cases of WBS patients with co-existing ASD features have been described and some authors have even suggested that ASD should be reconsidered as part of the WBS (7–18). Due to the low frequency of WBS and the even lower prevalence of

ASD features, most of studies focusing on the association both disorders are anecdotal reports or are based on small sample sizes. Methods and criteria to evaluate features of ASD also vary across studies and difficult drawing conclusions about the frequency of ASD features among WBS patients (19). However, a recent meta-analyses studying the presence of ASD features among several genetic syndromes concluded that the prevalence of ASD features among WBS individuals is approximately 12% (19). This represents a prevalence approximately ten times higher than among general population, a striking finding considering the typical neurocognitive profile of individuals with WBS.

To date, no genome-wide unbiased study has been carried out to look for second-hit genetic factors that could explain the presence of ASD features in some individuals with WBS. However, several explanations have been suggested. The mirror phenotypes observed in WBS and the 7q11.23 reciprocal duplication syndrome suggest that genes contained in the region are dosage sensitive and contribute to language development and behavior. In this regard, *GTF2I* and *GTF2IRD2*, the genes related to the cognitive phenotype, have been proposed as the main candidates. Given the association of the WBS locus with language and behavior, some authors have hypothesized that a trans-acting factor present in the remaining allele could act as a genetic modifying factor. Another possible explanation is a cis-acting mechanism in which breakpoints of the deletion would alter flanking genes, but given that nearly all of the patients described until the moment present the common deletion, this hypothesis seems improbable (15). Finally, a polygenic model where genes present in the WBS locus would interact with additional genetic factors creating a sensitized background and modify the phenotypic expression of the disorder has also been proposed (7,15). In this regard, two WBS-ASD patients, presenting hyperserotonemia, were described in a previous report (11,15). Both individuals were homozygous for the *short* (s) allele in the promoter of the serotonin transporter *SLC6A4* (5-HTTLPR), pointing towards the role of this locus as a possible modifying factor (15).

In the present work, we study second-hit modifying genetic factors in eight individuals with WBS and associated ASD. We have performed a genome-wide unbiased analysis by exome sequencing, which allowed us

to explore large and exonic Copy Number Variants (CNVs), as well as Single Nucleotide Variants (SNVs) in coding regions of the genome in a single assay. In addition, we have performed a detailed molecular characterization of the deletion and genotyped the previous candidate locus in the promoter of the serotonin transporter *SLC6A4*.

Materials and methods

Patient selection and deletion characterization

This study was conducted on eight WBS patients with associated ASD, consisting of four males and four females aged from 6 to 31 years old. All patients underwent an extensive clinical genetic examination and showed the distinct WBS facial dysmorphism. The clinical diagnosis was confirmed by molecular techniques and the size and parental origin of the deletion established using analysis of single and multiple-copy microsatellite as previously described (2). Based on the direct observation of a trained psychologist, a diagnosis of ASD was made and confirmed in seven of the patients by means of the Autism Diagnostic Interview-Revised (ADI-R). The remaining patient had a clinical diagnosis of ASD. Written informed consent was obtained from all parents or legal caregivers and the study was approved by the Clinical Research Ethics Committee of the centers involved (CEIC-Parc Salut Mar).

Exome capture and sequencing

Blood samples were obtained, and genomic DNA was extracted by the salting out method using the Puregene® DNA Purification Kit (Gentra Systems, Big Lake, MN, USA). Exomes were captured using the SureSelect Human All Exon V5 capture kit (Agilent, Santa Clara, CA, USA) and libraries were sequenced on an Illumina MiSeq platform. Paired-end sequences were obtained with a read length of 250 bp.

Mapping and calling

Reads were mapped to the hg19 version of the human genome using BWA following standard parameters. GATK's pipeline was applied for variant calling and SNV and indel calls were only considered if positions had a depth of coverage of at least 10× and a minimum of 20% of the

reads showed the variant (AB between 0.2 and 0.8). Annotation of variants was performed using ANNOVAR (http://www.openbioinformatics.org/annovar/), taking into account the variant frequency in control databases: dbSNP137 (http://www.ncbi.nlm.nih.gov/SNP/), ExAC (http://exac.broadinstitute.org/), Kaviar (http://db.systemsbiology.net/kaviar/) and an in-house database of 248 Spanish controls. The nature of the changes was assessed by PolyPhen and Condel (http://bg.upf.edu/fannsdb/) protein effect prediction algorithms.

Detection of second hit hemizygus SNVs in *GTF2I* and *GTF2IRD2*

Since SNVs in multi-copy genes, such as *GTF2I* and *GTF2IRD2*, are expected to have a lower allelic balance (AB) compared to those in single copy regions, we adapted mapping, variant calling and filtering to allow their detection. The pipeline followed for mapping and annotation of variants has been previously described in Chapter 3.

Genome-wide analysis of second hit CNVs

In order to detect additional pathogenic CNVs, we applied ExomeDepth, that uses read depth data to call CNVs and compared our samples to a matched aggregate reference set of 248 exomes captured and sequenced with the same protocol. CNVs were filtered based on their overlap with variants previously described in the Database of Genomic Variants (DGV), and DECIPHER (20)

Genome-wide analysis of rare variants

To look for rare second-hit SNVs, we selected only exonic variants with a functional effect and excluded variants previously described in the general population (Kaviar, ExAC and in-house database of Spanish controls) with a minor allele frequency (MAF)>0.01.

Sanger sequencing validation

To validate and perform segregation studies, we used Sanger sequencing by capillary electrophoresis (ABI PRISM 7900HT, Applied Biosystems, Foster City, CA, USA). Primers were designed with PRIMER 3 application

(http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/) and PCR reactions were carried following standard conditions.

Altered genes in SW

To study the contribution of epistatic effects, we established a list of altered genes in WBS patients by various mechanisms: transcriptional dysregulation, differentially methylation and direct GTF2I targets. We performed a systematic literature review and selected only high quality studies carried out in human subjects (21–24), obtaining a list of 2251 dysregulated genes.

Genome-wide analysis of ASD susceptibility loci

In order to detect common variants that could act as susceptibility variants, we focused on SNVs that had been previously associated to ASD by Genome-Wide Association Studies (GWAs). The list of associated SNPs was obtained from the Genome-Wide Repository of Associations between SNPs and Phenotypes (GRASP) database version 2.0 and all variants that contained the term “Asperger” or “Autism” under the phenotype category were selected (25). We then compared the allele frequency in our sample to that described in Non Finnish European population data form ExAC by Fisher’s Exact Test and calculated their estimated q-value and False Discovery Rate (FDR) values using the R package qvalue (26). Finally, to avoid population stratification, we compared ExAC allele frequencies to that reported in Spanish Variant Server, which contains data from 578 Spanish individuals (http://csvs.babelomics.org/).

SLC6A4 genotyping

The promoter polymorphism of the serotonin transporter *SLC6A4* (5-HTTLPR) was genotyped by PCR followed by agarose gel electrophoresis, using primers previously described (15).

Results

To look for second-hit genetic factors that could act as modifiers, we performed an unbiased genome-wide study of variation in coding regions of the genome following the strategy depicted in Figure 1.

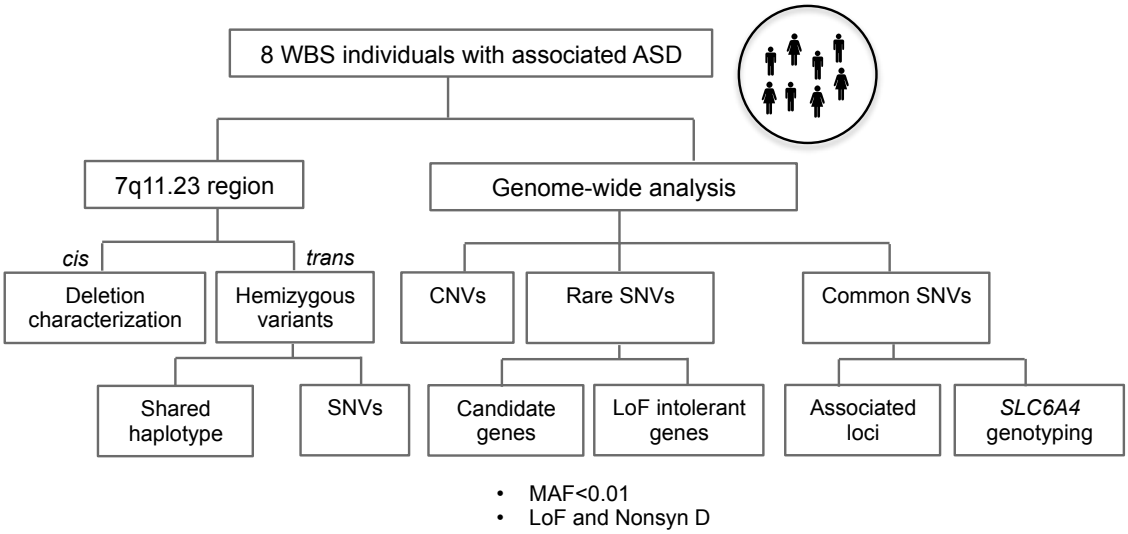


Figure 1. Strategy followed for the identification of second-hit genetic factors.

Patient description

The study has been carried out in eight Spanish WBS patients with a comorbid diagnosis of ASD, consisting of four males and four females aged from 6 to 31 years old. All patients underwent an extensive clinical genetic examination and the diagnosis of WBS was confirmed in all by molecular methods. All individuals had a clinical diagnosis of ASD, based on the direct observation of a trained psychologist, which was confirmed in seven individuals by means of the Autism Diagnostic Interview-Revised (ADI-R). The main clinical characteristics of patients are summarized in Table 1.

Patient	WBS1	WBS2	WBS3	WBS4	WBS5	WBS6	WBS7	WBS8
Birth year	1987	1985	2000	2006	2002	2008	2001	2010
Gender	F	M	M	M	F	M	F	F
Language	NV	V	V	NV	NV	NV	NV	V
Age at ASD clinical suspicion	4 y	10 y	6 y	5 y	12 y	5 y	3 y	5 y
ADI-R								
Developmental abnormalities evident <= 36 months	NA	4	1	5	3	4	4	5
Communication	NA	7	4	12	7	4	14	9
Reciprocal social interaction	NA	11	11	22	14	10	19	18
Repetitive behaviors and sterotyped patterns	NA	4	6	8	8	6	9	5
Attention-Deficit/Hyperactivity Disorder	Yes	Yes	Yes	ND	ND	Yes	ND	ND
Family history of neurodevelopmental or psychiatric disorders	No	No	Father' sister (ASD and profound ID)	No	No	Father died by suicide (no previous psychiatric disorders)	No	No
Age at WBS diagnosis	7 y	3 y	5 m	3 m	3 m	2 y	5 y	1 y
Cardiovascular disease	ND	SVAS	Coarctation of aorta	SVAS	SVAS	SVAS	SVAS	ND
Kidney abnormalities	ND	ND	ND	ND	ND	ND	ND	ND
Endocrine abnormalities	ND	ND	ND	ND	Early puberty	ND	Early puberty	Subclinical hypotiroidism

Table 1. Clinical characteristics of the eight Spanish WBS patients with associated ASD. V: verbal; NV: non verbal; SVAS: Supravalvular Aortic Stenosis; ND: not detected by standard testing.

Detailed characterization of the 7q11.23 deletion

To study the presence of cis-acting factors, we characterized all deletions at the molecular level by genotyping multiple polymorphic markers. The results showed that all rearrangements were *de novo*, being 7/8 of parental origin and one originated in the maternal allele (Table 2). Most of the patients (6/8) carried the common 1.55 Mb deletion mediated by blocks B, whereas 2/8 carried the larger 1.83 Mb deletion mediated by blocks A. In two patients, the common 1.55 Mb deletion had been mediated by an inversion in the transmitting progenitor. Since *GTF2IRD2* has been postulated as a possible modulator of the cognitive phenotype in WBS (27), we analyzed the deletion breakpoint to assess the number of functional copies of this gene (Figure 2).

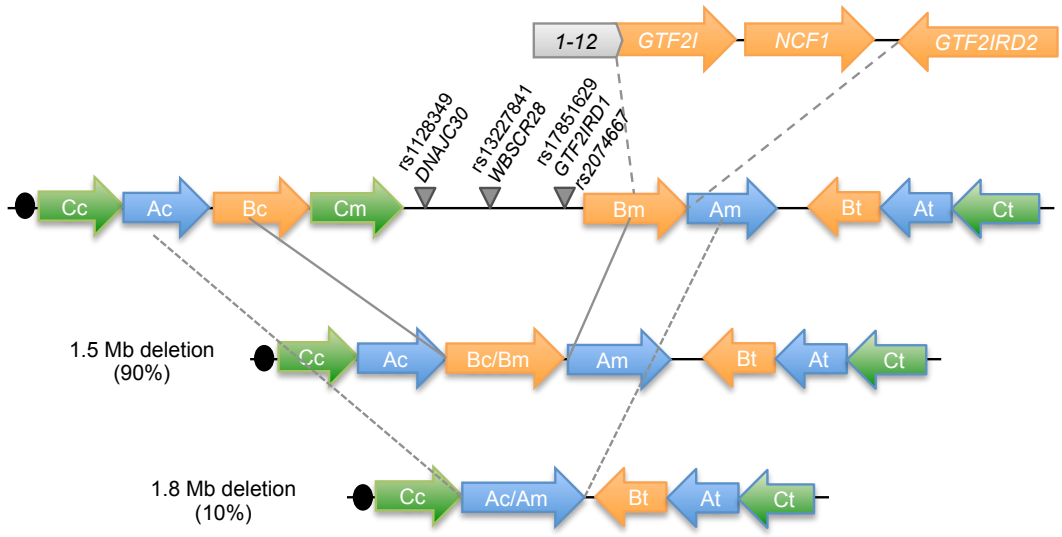


Figure 2. Schematic representation of the WBS locus, showing the two most common deletions and the gene content of the B block. The B block contains three genes: *GTF2I*, *NCF1* and *GTF2IRD2*. *GTF2I* and *NCF1* have a single functional copy, located at the medial B block, whereas *GTF2IRD2* has two functional copies, located at the medial and telomeric B blocks. The 1.55 Mb most common deletion is mediated by B blocks and results in a chimeric medial-centromeric block. The number of functional copies in this case depends on the deletion breakpoint. The 1.8 Mb deletion is mediated by A blocks, resulting in the loss of the medial and centromeric B blocks and an additional functional copy of *GTF2IRD2*.

As expected, the two deletions mediated by A blocks showed four copies of *GTF2IRD2*, since the centromeric and medial B blocks are lost, resulting in a centromeric (C), a medial (M) and two telomeric (T) (1C/1M/2T) copies of *GTF2IRD2*. The two patients with inversion-mediated deletions had a

1C/1M/3T genotype, originated from the loss of a medial copy and a gain of a telomeric copy of GTF2IRD2. Regarding the rest of the patients with 1.55 Mb deletions, in three the deletion breakpoint had occurred before *GTF2IRD2*, resulting in the loss of only a centromeric copy (1C/2M/2T), whereas in the remaining patient it had occurred within *GTF2IRD2,* creating a medial-centromeric chimeric copy of the gene (1Q/1M/2T).

	Size	**Inversion**	**Origin**	***GTF2IRD2***	***# functional copies***
WBS1	1,55	Yes	P	1C/1M/3T	4
WBS2	1,55	No	P	1C/2M/2T	4
WBS3	1,55	Yes	P	1C/1M/3T	4
WBS4	1,83	No	P	1C/1M/2T	3
WBS5	1,55	No	P	1C/2M/2T	4
WBS6	1,55	No	M	1Q/1M/2T	3+?
WBS7	1,55	No	P	1C/2M/2T	4
WBS8	1,83	No	P	1C/1M/2T	3

Table 2. Deletion characterization in the eight patients with WBS and ASD.

Second-hit hemizygous variants in the 7q11.23 region

To look for the presence of second hit trans-acting factors in the remaining 7q11.23 hemizygous allele, we performed a detailed analysis of the region and studied overrepresented common variants, shared haplotypes and rare deleterious SNVs. First, we looked for the presence of common variants or Single Nucleotide Polymorphisms (SNPs) that were overrepresented in our cohort. Allele frequencies of all previously described hemizygous variants (n=32) in our cohort were compared to those reported in ExAC. To avoid population stratification, variants with significantly different frequencies in ExAC and Spanish Variant Server were excluded. Due to our small sample, Fisher's exact test did not reveal any significant difference in allele frequency. However, we identified a nearly significant association (p=0.076), with two individuals (WBS2 and WBS6) carrying a SNP (rs12539160), that had been associated to ASD in a previous GWA study (28). The variant was a synonymous SNV lying in an exon boundary of *MLXIPL*, with a MAF of 0.06 in European population according to ExAC and of 0.03 in the Spanish Variant Server.

Taking advantage of the hemizygous state of SNPs located in the deleted single-copy region, we extracted phased haplotypes in the deleted single copy interval to study if a common haplotype was shared between individuals. Two linkage disequilibrium blocks were identified from rs1128349 to rs13227841 (*DNAJC30* to *WBSCR28*) and from rs17851629 to rs2074667 (located in *GTF2IRD1*) (Figure 2, Supplementary Table 1). Allele frequencies of the tag markers did not differ significantly from those in the general population and no common shared haplotype was identified in our cohort.

In addition, we looked for rare SNVs in the single-copy region of WBS. Only homozygous variants with a MAF<0.01 were selected. After filtering for exonic variants and excluding synonymous SNVs with no functional effect, we remained with two nonsynonymous variants in genes *MLXIPL* and *TBL2* predicted as tolerant by various protein effect prediction algorithms (SIFT, Polyphen and Condel). Besides hemizygous genes, we studied variation in *GTF2I,* considered the main candidate gene for the neurocognitive profile, and *GTF2IRD2*, one of the possible modulators. Since SNVs in duplicated genes, such as *GTF2I* and *GTF2IRD2*, are expected to have lower allelic balance (AB) ratios than those in single copy regions, we adapted mapping and variant calling settings to allow their detection, as detailed in the materials and methods section. After filtering variants with a MAF<0.01 we did not identify any variant.

Genome-Wide Analysis of second-hit variants

Copy Number Variants

Since large CNVs acting as two-hit factors have been described in patients with neurodevelopmental disorders carriers of CNVs of incomplete penetrance, we studied the presence of additional rearrangements that could explain the presence of autistic symptoms in our cohort. A part of the WBS deletion, we observed an average of 25 additional CNVs per patient ranging for 170 bp to 334 kb. None of the additional CNVs overlapped with known genomic disorders or was previously associated to neurodevelopmental disorders and all were present in the general population. All CNVs containing genes were intersected with a list of candidate loci previously associated to ASD (SFARI gene list), resulting in two CNVs overlapping with candidate genes *SIK1* and *DUPS22*. The CNV affecting *SIK1*, found in patient WBS3,

was a partial duplication affecting exons 1 and 2, whereas a complete heterozygous deletion containing *DUSP22* was found in patient WBS7. Both rearrangements overlapped with previous CNVs described in the general population (29,30).

Rare Single Nucleotide variants

We next analyzed the presence of second-hit rare SNVs that could act as susceptibility factors to ASD. Since second-hit variants are not expected to be the sole cause of ASD in individuals with WBS, they might be present in the general population and inherited from unaffected progenitors. Moreover, given the high degree of genetic heterogeneity of ASD, they could affect hundreds of genes. Taking into account the prevalence of ASD in individuals with WBS and the large number of genes involved, we reasoned that the frequency of each individual variant should be relatively rare and, consequently, set our MAF threshold at ≤0.01. To prioritize SNVs that could act as two-hit risk factors, we selected deleterious variation, defined as LoF and nonsynonymous damaging variants predicted by both SIFT and Polyphen, affecting candidate genes from SFARI (n=792) (31). In addition, we focused on deleterious variants present in functionally constrained genes, defined as those with a probability of being LoF intolerant (pLI)>0.9 (32).

Regarding variants in candidate genes, we detected a total of 39 rare deleterious SNVs, found in 35 genes. Six of the identified variants were LoF (Table 3), whereas the rest were missense variants predicted as damaging (Supplementary Table 2). It is interesting that five of the genes with LoF variants in our cohort (*PYHIN1*, *UBR5*, *AGAP1, CC2D1A* and *USP45)* were proposed as candidate loci for ASD due to the presence of *de novo* LoF variants. In addition, two of them (*AGAP1* and *UBR5*) are functionally constrained genes with a very low burden of LoF mutations in the general population. In addition, we found a novel frameshift variant in *PIK3CG*, which results in a prematurely truncated protein, containing only exon one and part of exon two. *PIK3CG* encodes for a catalytic subunit of phosphoinositide 3-kinase (PI3K) part of the PI3K/AKT/mTOR intracellular signaling pathway, which regulates multiple cellular functions and has been extensively associated to ASD (33–35).

Individual	Gene	Variant	pLI	MAF	Evidence	Ref
WBS_4	*CC2D1A*	c.1357-2A>C	0	4,4E-04	1,2	(36,37)
WBS_5	*PYHIN1*	p.R373X	0	2,8E-04	*2*	(38)
WBS_5	*UBR5*	p.R633X	1	0	*2*	(38–41)
WBS_6	*AGAP1*	c.1051-2A>T	1	2,0E-04	1,2	(37,42–44)
WBS_7	*USP45*	p.E220X	0	6,5E-04	1,2	(37,45)
WBS_8	*PIK3CG*	p.12_13del	0,56	0	3,4	(33–35)

Table 3. LoF mutations identified in candidate genes in five different patients.

Among candidate genes harboring deleterious SNVs, we identified four recurrently mutated loci: *HLA-A*, *SDK1*, *PCDH15* and *TSC2*. *HLA-A*, *SDK1* and *PCDH15* had been previously described as risk loci in several case-control association studies but, to our knowledge, no rare variants have been described in individuals with ASD (46–50). Clinical examination of the two individuals carrying rare variants in *TSC2*, WBS_1 and WBS_7, did not reveal any suggestive sign of tuberous sclerosis, although rare variants in *TSC2* have also been associated to non-syndromic ASD (51,52).

Previous studies have shown that genes predisposing to ASD carry a low burden of disrupting mutations in the general population, since they usually result in phenotypes of reduced fecundity (53). Therefore, we analyzed deleterious variation in highly constrained genes, defined as those with a pLI > 0.9 (31). Excluding candidate genes, we detected 68 SNVs in 64 genes, including five LoF mutations (Table 4) and 63 missense damaging variants (Supplementary Table 3).

Individual	Gene	Variant	MAF	Inheritance
WBS3	*PRR12*	p.P482fs	NA	NA
WBS5	*SEC24C*	c.2682+2A>G	NA	NA
WBS7	*CXXC1*	p.S222fs	NA	NA
WBS8	*MED26*	c.147+2A>G	87/154602	NA
WBS8	*EPHB1*	p.A433fs	NA	Maternal

Table 4. LoF mutations identified in functionally constrained genes (pLI>0.9).

Interestingly, all genes with LoF in our cohort, but very few in the general population, are brain expressed genes and some have been previously

associated with psychiatric and neurodevelopmental disorders. Three of the LoF variants affecting *PRR12*, *SEC24C*, *CXXC1* and *EPHB1*, were novel and had not been described in 77,781 individuals from Kaviar database. *PRR12* was previously found to be disrupted in a girl with ID and psychiatric alterations (54). *SEC24C* is involved in vesicle trafficking and in the transport of the serotonin transporter and has been associated to bipolar disorder by topological analysis of protein-protein interaction (PPI) networks (55,56). *CXXC1* encodes a transcriptional activator that binds specifically to non-methylated CpG and interacts with *SETD1A*, a chromatin regulator associated to schizophrenia and developmental disorders (40,57). *EPHB1* belongs to the ephrin family, which play an important role in axon guidance and dendritic spine maduration, and has been described as a schizophrenia risk loci in a GWAs study in Japanese population (58,59). Moreover, *EPHB1* was recurrently mutated in our cohort, since a second patient (WBS7) carried an additional rare missense damaging variant. Finally, the same patient carrying a maternally inherited LoF mutation in *EPHB1*, WBS8, also carried a reported mutation in *MED26*, a transcription regulator factor involved in neuronal gene expression (60).

Epistatic effects of rare variants

To look for second hits disrupting the same pathway that could cause a more severe phenotype by an epistatic effect, we intersected a manually curated list of genes altered in WBS WITH our previously selected candidate and functionally constrained genes (Figure 3). The results showed that four genes altered in WBS, that were previously described as ASD candidate genes and functionally constrained, harbored rare deleterious mutations in our cohort. These included *AGAP1*, with a LoF mutation in our cohort. Interestingly, *EPHB1*, which is recurrently mutated in our cohort, was also altered in WBS.

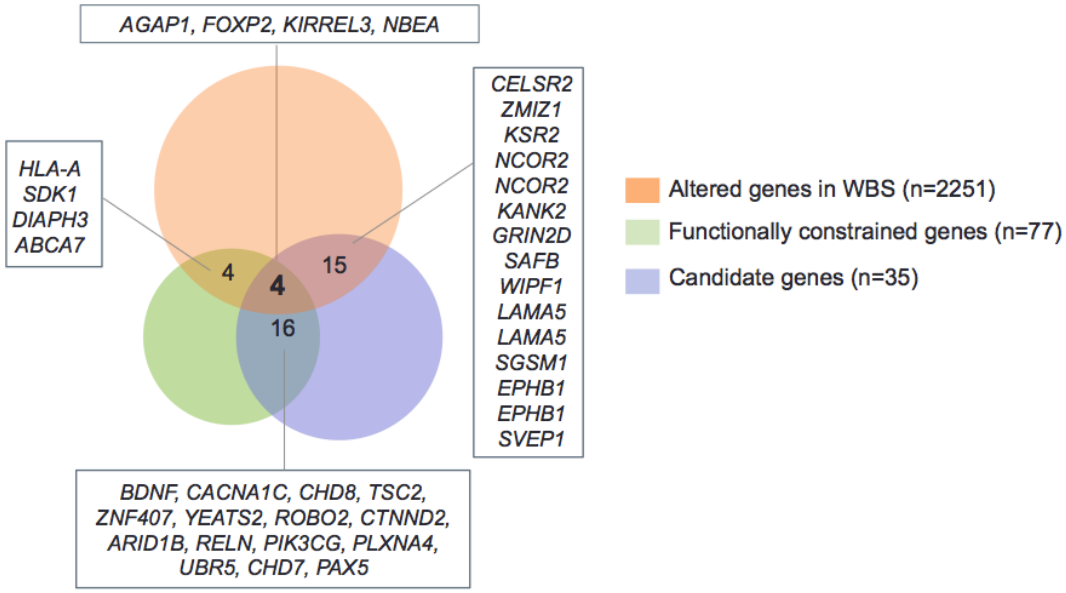


Figure 3. Overlap between candidate and functionally constrained genes mutated in our cohort and genes altered in WBS.

Increased burden of deleterious mutations in females

Since ASD are four times more prevalent in males than females, it has been suggested that females have a higher risk threshold and require a higher genetic load to develop the disorder (61). In fact, several studies have shown that *de novo* LoF variants are more frequent in affected females than affected males and that there is a maternal bias in the inherited of truncating variants and second-hit CNVs (38,39,62,63).

To examine if females in our cohort carried a higher genetic burden, we compared the frequency of rare deleterious variants in ASD candidate genes between affected males (n=4) and affected females (n=4) (Figure 4A). The results showed a statistically significant increase (p=0.016) in the average number of mutations per patient in females ($\bar{x} = 6$, SD=1,7) compared to males ($\bar{x} = 3$, SD=1,2). This effect was not due to an increased prevalence of rare variants between genders, since the global frequency of rare variants was not significant different between both genders (p=0.51, Figure4B). Finally, we compared the frequency of deleterious variants in functionally constrained genes between both sexes and found a nearly significant association (p=0.064, Figure 4C)), that was reduced when candidate genes (n=19) were excluded from the analysis (p=0.15, Figure 4D).

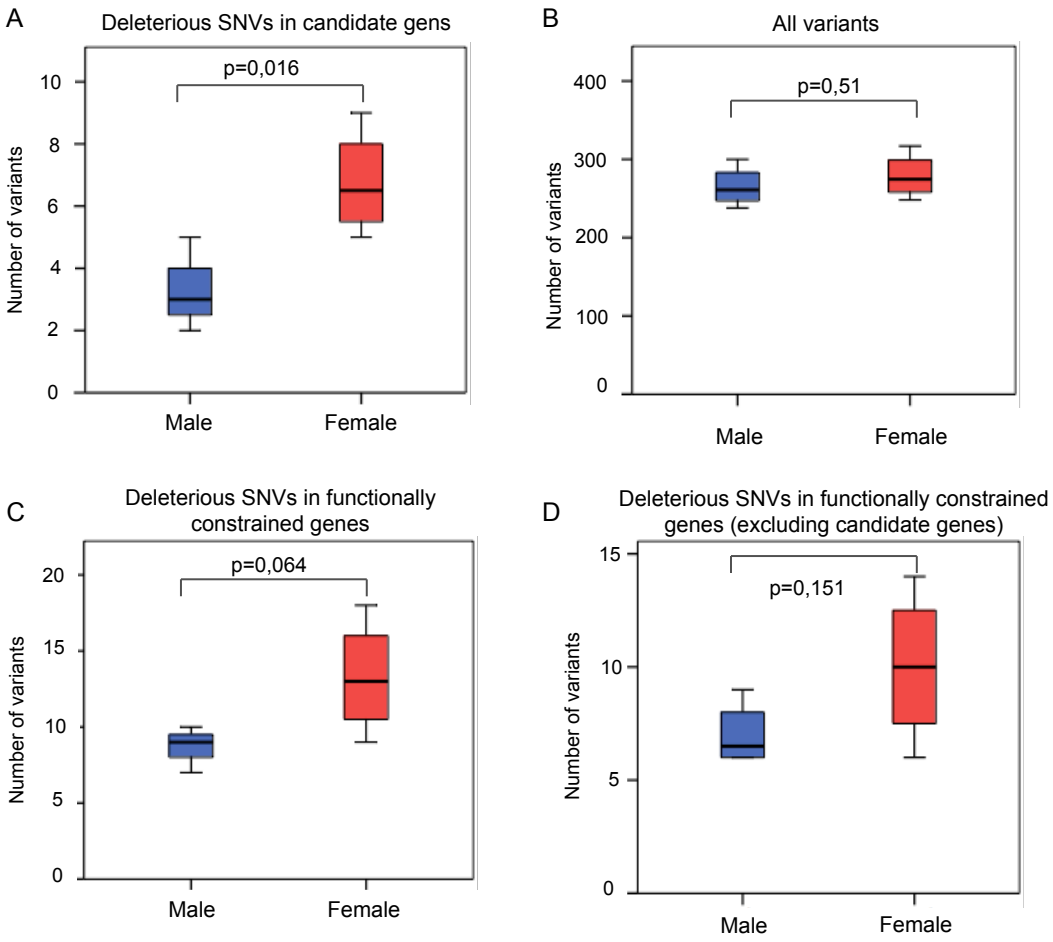


Figure 4. A) The number of rare (MAF<0.01) deleterious mutations per individual in candidate genes is compared in females (n=4) and males (n=4). B) The number of rare variants (MAF<0.01) per individual is compared in females (n=4) and males (n=4). C) The burden of rare variants (MAF<0.01) per individual in functionally constrained genes (pLI>0.9) is compared in females (n=4) and males (n=4), including candidate genes. D) The number of rare variants (MAF<0.01) per individual in functionally constrained genes (pLI>0.9), excluding previously reported candidate genes, is compared in females (n=4) and males (n=4). Statistical significance in all tests was calculated using a two-sided student's t-test, excluding SNVs present in the X and Y chromosomes from the analysis.

Genome-wide analysis of ASD susceptibility loci

Finally, to study the role of common susceptibility variants, we have performed a genome-wide study of all SNPs previously associated to ASD by GWAs. To this end, we compared the frequency of previously reported associated markers in our cohort to that of non Finnish European population in ExAC. A total of 645 single-nucleotide markers located in coding regions

were analyzed, of which only 13 obtained a significant p.value ($p<0.05$) and none passed FDR correction. For ten of the significantly overrepresented SNVs in our cohort, the overrepresented allele corresponded to the risk allele, whereas for the remaining, it was the protector allele (Supplementary Table 4). Three of the markers had been previously described as Expression Quantitative Trait Loci (eQTL) in brain, with rs2275477 associated with increased expression of *OSCP1* and rs4823086 and rs5749088 of transcript RP1.130H16.16, located in *CCDC157* (64). Interestingly, we identified a nominally significant increase in the frequency of a marker located in *PCDH15* (rs2135720) (46). This gene was recurrently mutated in our study of candidate genes, with two deleterious mutations in two individuals (WBS_1 and WBS_8) who where not carriers of the previously mentioned risk allele.

SLC6A4 genotyping

Since *SLC6A4* had been proposed as a modifying locus for the phenotypic outcome observed in WBS in two WBS patients with autistic symptoms and hyperserotonemia, we genotyped the known polymorphism located in the promoter, as previously described (15). In our cohort, the genotype frequencies were similar to those found in population of European origin, with 3/8 individuals homozygous for the major long variant (*l*), 4/8 presenting an heterozygous genotype with both l and short (*s*) variant, and 1/8 individual being homozygous for the s allele.

Discussion

In this study we have performed a comprehensive analysis of second-hit genetic factors in eight patients presenting WBS and associated ASD by exome sequencing, which allows the detection of both SNVs and CNVs at a genome-wide level. To our knowledge, this is the first study exploring genetic modifying factors in WBS with associated ASD in an unbiased manner. To date, the only locus suggested as a modifying factor of the WBS neurocognitive profile is the serotonin transporter SLC6A4. A previous study by Tordjman et al, reported two WBS individuals with associated ASD and hyperserotonemia, both homozygous carriers of the ss polymorphism of the promoter of the serotonin transporter (5-HTTLPR). (15). To confirm this association, we genotyped this locus in our cohort, showing that the genotype frequencies were similar to those found in population of European origin.

Since hyperserotonemia is a frequent finding among individuals with ASD, the question remains if it plays a primary role in the associated ASD features seen in WBS patients or if it appears as a secondary finding.

Similar to previous reports, the detailed molecular characterization of the deletions showed no atypical rearrangement, excluding the alteration of flanking genes by atypical rearrangements as a possible explanation. Moreover, the detailed characterization of the number and type of *GTF2IRD2* copies showed no differences with those previously reported in WBS patients with the typical neurocognitive profile (2). However, in seven of eight patients, the deletion had originated in the paternal allele, suggesting a role of epigenetic control mechanisms. In fact, it has been shown that the expression of *GTF2I* is lower in WBS patients with deletions of paternal origin, and it has been hypothesized *GTF2I* is paternally imprinted (65,66). Although several cases of patients with WBS and associated ASD have been reported in the literature, we could not find any study evaluating parental origin of the deletion. Taking into account the prevalence of ASD among WBS patients and the fact that deletions occur with equal frequency in the maternal and paternal allele, parental origin cannot be the sole factor influencing ASD risk. Molecular characterization of the parental origin of the deletion in additional patients presenting ASD features can shed more light on this matter.

Second hit hemizygous variants in the 7q11.23 region disrupting the same pathway could also act as modifying factors with an epistatic effect. The detailed analysis of SNVs did not reveal any common haplotype or any rare variant in the coding regions of *GTF2I* or *GTF2IRD2,* the main candidate genes for the neurocognitive profile observed in WBS. However, the analysis of overrepresented common variants revealed that two individuals (WBS2 and WBS6) carried a SNP (rs12539160) in hemizygous state previously associated to ASD. This variant is a synonymous mutation located in *MLXIPL*, a candidate gene for the lipid abnormalities observed in WBS patients (67). Although this marker may not be responsible for the increased risk of ASD, it may be in linkage disequilibrium with a causative SNP in a noncoding region, not covered by exome sequencing. Additional studies examining variation in noncoding regions in a larger cohort of individuals with WBS and ASD features are necessary to confirm this finding.

Since deletion characterization and analysis of hemizygous variation did not reveal any clear second-hit factor that could explain the presence of ASD in these individuals, we studied second-hit variation at a genome-wide level. An advantage of exome sequencing is its potential to detect both CNVs and SNVs, as shown by the validation of all WBS deletions. However, no additional second-hit CNV was identified in any of the patients. Likewise, a previous study evaluating second-hit CNVs in genomic disorders showed that second-site variants are more frequent in disorders of variable expressivity than in syndromic entities, such as WBS (63). In fact in this study, second-site variants were only detected in 5% of patients with WBS, suggesting that additional large CNVs may be incompatible with life. Our results support this hypothesis, since no rare large additional CNV was identified in any of the patients.

Besides CNV, we examined the effect of rare deleterious SNVs, focusing on ASD candidate genes and functionally constrained genes, as they have been shown to be enriched in ASD susceptibility genes (53). By this approach, we have uncovered several candidate genes harboring LoF mutations in our cohort, as well as LoF mutations in brain-expressed genes with very few LoF mutations in the general population. Interestingly, two of the candidate genes with LoF mutations in our cohort, *USP45* and *PHYN1*, harbor LoF mutations in the general population, but are recurrently *de novo* mutated in ASD patients and not in their unaffected siblings (37–41,45). This suggests that *de novo* deleterious mutations in these genes could have a milder effect and would require a sensitized background to manifest their effects. In our case, the WBS deletion would act as major genetic lesion, whereas in ASD patients without a major hit, it could act as part of an oligogenic model. It is also noticeable that all genes with LoF variants in our cohort selected on the basis of being intolerant to deleterious mutations (*PRR12*, *SEC24C*, *CXXC1*, *MED26* and *EPHB1*) are brain-expressed genes, with some cases even being associated to various neuropsychiatric disorders. Moreover, two of the genes harboring LoF mutations, *EPHB1* and *AGAP1*, have been shown to be altered in WBS, suggesting a possible epistatic effect. *AGAP1* was found to be hypermethylated in WBS patients compared to individuals with the reciprocal duplication, which is associated to a phenotype of language impairment, anxiety and increased risk of ASD and schizophrenia (24,68–70). *EPHB1* was underexpressed in WBS patients and individuals with atypical

deletions and low IQ, suggesting a possible epistatic mechanism in which the frameshift mutation described in our patient would result in an even lower expression, possibly mediated by non-sense mediated decay (21). Taking together, these evidences suggest that inherited rare LoF in brain-expressed genes could play a major role in the susceptibility to ASD in these patients by both additive and epistatic effects.

Interestingly, we show that the number of rare deleterious variants in candidate genes is higher in females with WBS and associated ASD than males, in agreement with previous studies suggesting that females require a higher burden of damaging variants to develop ASD. We also observed a tendency for variants contained in highly functionally constrained genes, which lost significance when excluding reported candidate genes. A higher burden of second-site deletions disrupting ASD associated genes has also been observed in female patients carrying susceptibility variants at 16p11.2 (71). Although our small sample size requires caution interpreting the results, the same effect was not observed when considering all rare variants. Further studies assessing the difference in prevalence and severity of ASD features between genders in individuals with WBS will help to clarify if a higher burden of deleterious variation is also seen in females carrying fully penetrant syndromic variants, such as the WBS deletion.

Finally, to assess if common variation contributed to the different behavioral phenotype observed in these patients, we studied if any variant previously associated to ASD was overrepresented in our cohort. Although none of the variants passed FDR correction, three of the analyzed markers with nominal p-values had been previously described as an eQTL in brain, suggesting that these or other variants in LD could have a direct functional effect. However, this analysis has two clear limitations. First, our small sample size and the expected low effect of common variation pose a great limitation for statistical power. Second, most of the interrogated variation in previous GWAs studies resides in noncoding regions not covered by exome sequencing and, therefore, only 645 out of a total of 23847 previously associated SNPs could be evaluated in this study. Moreover, this limitation did not allow us to calculate polygenic risk scores, as previously done by various studies evaluating the effect of common variation on schizophrenia risk in patients with 22q11.2 deletions or susceptibility CNVs (72,73). The results of these

two studies suggest that common variation en masse could affect disease susceptibility and that a sensitized background may act as a susceptibility factor for additional neuropsychiatric phenotypes. Therefore, it would be of great interest to see if polygenic risk scores are also increased in WBS patients with accompanying ASD compared to WBS individuals showing the typical neurocognitive profile. This study would require whole-genome genotyping of associated markers, as well as in deep studies of the effect of combined common variation on ASD risk, since current polygenic scores for ASD explain only a small proportion of the variance (<1%), representing only about a third part of that explained for schizophrenia (46,74).

Although our work represents the first genome-wide study carried out in WBS patients with associated ASD, it presents several limitations. First, our sample size is small, even though this fact is explained by the low prevalence of WBS and the even lower prevalence of associated ASD. Second, exome sequencing does not interrogate noncoding variation. However, our unbiased analysis of second-hit factors suggests that rare deleterious coding variants in brain expressed genes that are functionally constrained could play a role in the phenotypic variability observed in patients, both by additive and epistatic effects. The higher burden of deleterious variants in females suggests that the female protective effect may still play a role in the phenotypic variability observed in genomic disorders of full penetrance, although further studies are needed to confirm that hypothesis. Finally, the study of second-hit SNVs in disorders of full penetrance, but variable expressivity, can help uncover new genes with milder effects that manifest only in a sensitized background, such as in the presence of a major genetic lesion.

References

1. Strømme P, Bjørnstad PG, Ramstad K. Prevalence estimation of Williams syndrome. J Child Neurol. 2002 Apr;17(4):269–71.

2. Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome deletions. Am J Hum Genet. 2003 Jul;73(1):131–51.

3. Howald C, Merla G, Digilio MC, Amenta S, Lyle R, Deutsch S, et al. Two high throughput technologies to detect segmental aneuploidies identify new Williams-Beuren syndrome patients with atypical deletions. J Med Genet. 2006 Mar;43(3):266–73.

4. Merla G, Brunetti-Pierri N, Micale L, Fusco C. Copy number variants at Williams-Beuren syndrome 7q11.23 region. Hum Genet. 2010 Jul;128(1):3–26.

5. Jones W, Bellugi U, Lai Z, Chiles M, Reilly J, Lincoln A, et al. II. Hypersociability in Williams Syndrome. J Cogn Neurosci. 2000;12 Suppl 1:30–46.

6. Asada K, Itakura S. Social phenotypes of autism spectrum disorders and williams syndrome: similarities and differences. Front Psychol. 2012;3:247.

7. Tordjman S, Anderson GM, Botbol M, Toutain A, Sarda P, Carlier M, et al. Autistic disorder in patients with Williams-Beuren syndrome: a reconsideration of the Williams-Beuren syndrome phenotype. PLoS One. 2012;7(3):e30778.

8. Gosch A, Pankau R. "Autistic" behavior in two children with Williams-Beuren syndrome. Am J Med Genet. 1994 Oct 15;53(1):83–4.

9. Herguner S, Mukaddes NM. Autism and Williams syndrome: a case report. World J Biol Psychiatry. 2006;7(3):186–8.

10. Reiss AL, Feinstein C, Rosenbaum KN, Borengasser-Caruso MA, Goldsmith BM. Autism associated with Williams syndrome. J Pediatr. 1985 Feb;106(2):247–9.

11. August GJ, Realmuto GM. Williams syndrome: serotonin's association with developmental disabilities. J Autism Dev Disord. 1989 Mar;19(1):137–41.

12. Gillberg C, Rasmussen P. Brief report: four case histories and a literature review of Williams syndrome and autistic behavior. J Autism Dev Disord. 1994 Jun;24(3):381–93.

13. Edelmann L, Prosnitz A, Pardo S, Bhatt J, Cohen N, Lauriat T, et al. An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism. J Med Genet. 2007 Feb;44(2):136–43.

14. Kilinçaslan A, Tanidir C, Tutkunkardaş MD, Mukaddes NM. Asperger's disorder and Williams syndrome: a case report. Turk J Pediatr. 53(3):352–5.

15. Tordjman S, Anderson GM, Cohen D, Kermarrec S, Carlier M, Touitou Y, et al. Presence of autism, hyperserotonemia, and severe expressive language impairment in Williams-Beuren syndrome. Mol Autism. 2013;4(1):29.

16. Klein-Tasman BP, Mervis CB, Lord C, Phillips KD. Socio-communicative deficits in young children with Williams syndrome: performance on the Autism Diagnostic Observation Schedule. Child Neuropsychol. 2007 Sep;13(5):444–67.

17. Klein-Tasman BP, Phillips KD, Lord C, Mervis CB, Gallo FJ. Overlap with the

autism spectrum in young children with Williams syndrome. J Dev Behav Pediatr. 2009 Aug;30(4):289–99.

18. Lincoln AJ, Searcy YM, Jones W, Lord C. Social interaction behaviors discriminate young children with autism and Williams syndrome. J Am Acad Child Adolesc Psychiatry. 2007 Mar;46(3):323–31.

19. Richards C, Jones C, Groves L, Moss J, Oliver C. Prevalence of autism spectrum disorder phenomenology in genetic disorders: a systematic review and meta-analysis. The lancet Psychiatry. 2015 Oct;2(10):909–16.

20. Firth H V., Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet. 2009;84(4):524–33.

21. Antonell A, Del Campo M, Magano LF, Kaufmann L, de la Iglesia JM, Gallastegui F, et al. Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. J Med Genet. 2010 May;47(5):312–20.

22. Henrichsen CN, Csárdi G, Zabot M-T, Fusco C, Bergmann S, Merla G, et al. Using transcription modules to identify expression clusters perturbed in Williams-Beuren syndrome. PLoS Comput Biol. 2011;7(1):e1001054.

23. Adamo A, Atashpaz S, Germain P-L, Zanella M, D'Agostino G, Albertin V, et al. 7q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. Nat Genet. 2015 Feb;47(2):132–41.

24. Strong E, Butcher DT, Singhania R, Mervis CB, Morris CA, De Carvalho D, et al. Symmetrical Dose-Dependent DNA-Methylation Profiles in Children with Deletion or Duplication of 7q11.23. Am J Hum Genet. 2015 Aug 6;97(2):216–27.

25. Eicher JD, Landowski C, Stackhouse B, Sloan A, Chen W, Jensen N, et al. GRASP v2.0: an update on the Genome-Wide Repository of Associations between SNPs and phenotypes. Nucleic Acids Res. 2015 Jan;43(Database issue):D799–804.

26. Bass, J., Dabney, A., Robinson D. qvalue: Q-value estimation for false discovery rate control. 2015.

27. Porter MA, Dobson-Stone C, Kwok JBJ, Schofield PR, Beckett W, Tassabehji M. A role for transcription factor GTF2IRD2 in executive function in Williams-Beuren syndrome. PLoS One. 2012;7(10):e47457.

28. Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, et al. A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet. 2010;19(20):4072–82.

29. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat Genet. 2011 Sep;43(9):838–46.

30. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, et al.

Genetic and functional analyses of SHANK2 mutations suggest a multiple hit model of autism spectrum disorders. PLoS Genet. 2012 Feb;8(2):e1002521.

31. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet. 2014 Sep;46(9):944–50.

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

33. Serajee FJ, Nabi R, Zhong H, Mahbubul Huq AHM. Association of INPP1, PIK3CG, and TSC2 gene variants with autistic disorder: implications for phosphatidylinositol signalling in autism. J Med Genet. 2003 Nov;40(11):e119.

34. Cuscó I, Medrano A, Gener B, Vilardell M, Gallastegui F, Villa O, et al. Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. Hum Mol Genet. 2009 May 15;18(10):1795–804.

35. Chen J, Alberts I, Li X. Dysregulation of the IGF-I/PI3K/AKT/mTOR signaling pathway in autism spectrum disorders. Int J Dev Neurosci. 2014 Jun;35:35–41.

36. Manzini MC, Xiong L, Shaheen R, Tambunan DE, Di Costanzo S, Mitisalis V, et al. CC2D1A regulates human intellectual and social function as well as NF-κB signaling homeostasis. Cell Rep. 2014 Aug 7;8(3):647–55.

37. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014 Nov 13;515(7526):216–21.

38. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015 Jun;47(6):582–8.

39. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014 Oct 29;515(7526):209–15.

40. Singh T, Kurki MI, Curtis D, Purcell SM, Crooks L, McRae J, et al. Rare loss-of-function variants in SETD1A are associated with schizophrenia and developmental disorders. Nat Neurosci. 2016 Apr;19(4):571–7.

41. Takata A, Ionita-Laza I, Gogos JA, Xu B, Karayiorgou M. De Novo Synonymous Mutations in Regulatory Elements Contribute to the Genetic Etiology of Autism and Schizophrenia. Neuron. 2016 Mar 2;89(5):940–7.

42. Wassink TH, Piven J, Vieland VJ, Jenkins L, Frantz R, Bartlett CW, et al. Evaluation of the chromosome 2q37.3 gene CENTG2 as an autism susceptibility gene. Am J Med Genet B Neuropsychiatr Genet. 2005 Jul 5;136B(1):36–44.

43. Bergen SE, O'Dushlaine CT, Ripke S, Lee PH, Ruderfer DM, Akterin S, et al. Genome-wide association study in a Swedish population yields support for

greater CNV and MHC involvement in schizophrenia compared with bipolar disorder. Mol Psychiatry. 2012 Sep;17(9):880–6.

44. Schizophrenia Psychiatric Genome-Wide Association Study (GWAS) Consortium. Genome-wide association study identifies five new schizophrenia loci. Nat Genet. 2011 Oct;43(10):969–76.

45. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015 Sep 23;87(6):1215–33.

46. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, et al. Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Hum Mol Genet. 2012;21(21):4781–92.

47. Connolly JJ, Glessner JT, Hakonarson H. A genome-wide association study of autism incorporating autism diagnostic interview-revised, autism diagnostic observation schedule, and social responsiveness scale. Child Dev. 84(1):17–33.

48. Torres AR, Sweeten TL, Cutler A, Bedke BJ, Fillmore M, Stubbs EG, et al. The association and linkage of the HLA-A2 class I allele with autism. Hum Immunol. 67(4-5):346–51.

49. Al-Hakbany M, Awadallah S, Al-Ayadhi L. The Relationship of HLA Class I and II Alleles and Haplotypes with Autism: A Case Control Study. Autism Res Treat. 2014;2014:242048.

50. Nudel R, Simpson NH, Baird G, O'Hare A, Conti-Ramsden G, Bolton PF, et al. Associations of HLA alleles with specific language impairment. J Neurodev Disord. 2014;6(1):1.

51. Kelleher RJ, Geigenmüller U, Hovhannisyan H, Trautman E, Pinard R, Rathmell B, et al. High-throughput sequencing of mGluR signaling pathway genes reveals enrichment of rare variants in autism. PLoS One. 2012;7(4):e35003.

52. Schaaf CP, Sabo A, Sakai Y, Crosby J, Muzny D, Hawes A, et al. Oligogenic heterozygosity in individuals with high-functioning autism spectrum disorders. Hum Mol Genet. 2011 Sep 1;20(17):3366–75.

53. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee Y-H, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A. 2015;112(41):E5600–7.

54. Córdova-Fletes C, Domínguez MG, Delint-Ramirez I, Martínez-Rodríguez HG, Rivas-Estilla AM, Barros-Núñez P, et al. A de novo t(10;19)(q22.3;q13.33) leads to ZMIZ1/PRR12 reciprocal fusion transcripts in a girl with intellectual disability and neuropsychiatric alterations. Neurogenetics. 2015 Oct;16(4):287–98.

55. Lee S-A, Tsao TT-H, Yang K-C, Lin H, Kuo Y-L, Hsu C-H, et al. Construction and analysis of the protein-protein interaction networks for schizophrenia, bipolar disorder, and major depression. BMC Bioinformatics. 2011;12 Suppl 1:S20.

56. Montgomery TR, Steinkellner T, Sucic S, Koban F, Schüchner S, Ogris E, et al. Axonal targeting of the serotonin transporter in cultured rat dorsal raphe neurons is specified by SEC24C-dependent export from the endoplasmic reticulum. J Neurosci. 2014 Apr 30;34(18):6344–51.

57. Tate CM, Lee J-H, Skalnik DG. CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. FEBS J. 2010 Jan;277(1):210–23.

58. Egea J, Klein R. Bidirectional Eph-ephrin signaling during axon guidance. Trends Cell Biol. 2007 May;17(5):230–8.

59. Ikeda M, Aleksic B, Kinoshita Y, Okochi T, Kawashima K, Kushima I, et al. Genome-wide association study of schizophrenia in a Japanese population. Biol Psychiatry. 2011 Mar 1;69(5):472–8.

60. Ding N, Tomomori-Sato C, Sato S, Conaway RC, Conaway JW, Boyer TG. MED19 and MED26 are synergistic functional targets of the RE1 silencing transcription factor in epigenetic silencing of neuronal gene expression. J Biol Chem. 2009 Jan 30;284(5):2648–56.

61. Robinson EB, Lichtenstein P, Anckarsäter H, Happé F, Ronald A. Examining and interpreting the female protective effect against autistic behavior. Proc Natl Acad Sci U S A. 2013;110(13):5258–62.

62. Jacquemont S, Coe BP, Hersch M, Duyzend MH, Krumm N, Bergmann S, et al. A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. Am J Hum Genet. 2014;94(3):415–25.

63. Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. N Engl J Med. 2012;367(14):1321–31.

64. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun;45(6):580–5.

65. Collette JC, Chen X-N, Mills DL, Galaburda AM, Reiss AL, Bellugi U, et al. William's syndrome: gene expression is related to parental origin and regional coordinate control. J Hum Genet. 2009 Apr;54(4):193–8.

66. Pérez Jurado LA, Peoples R, Kaplan P, Hamel BC, Francke U. Molecular definition of the chromosome 7 deletion in Williams syndrome and parent-of-origin effects on growth. Am J Hum Genet. 1996 Oct;59(4):781–92.

67. Palacios-Verdú MG, Segura-Puimedon M, Borralleras C, Flores R, Del Campo M, Campuzano V, et al. Metabolic abnormalities in Williams-Beuren syndrome. J Med Genet. 2015 Apr;52(4):248–55.

68. Mulle JG, Pulver AE, McGrath JA, Wolyniec PS, Dodd AF, Cutler DJ, et al. Reciprocal duplication of the Williams-Beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. Biol Psychiatry. 2014 Mar 1;75(5):371–7.

69. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca

D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011 Jun 9;70(5):863–85.

70. Mervis CB, Klein-Tasman BP, Huffman MJ, Velleman SL, Pitts CH, Henderson DR, et al. Children with 7q11.23 duplication syndrome: Psychological characteristics. Am J Med Genet Part A. 2015;n/a – n/a.

71. Duyzend MH, Nuttle X, Coe BP, Baker C, Nickerson DA, Bernier R, et al. Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. Am J Hum Genet. 2016 Jan 7;98(1):45–57.

72. Tansey KE, Rees E, Linden DE, Ripke S, Chambert KD, Moran JL, et al. Common alleles contribute to schizophrenia in CNV carriers. Mol Psychiatry. 2015 Sep 22;

73. Merico D, Zarrei M, Costain G, Ogura L, Alipanahi B, Gazzellone MJ, et al. Whole-Genome Sequencing Suggests Schizophrenia Risk Mechanisms in Humans with 22q11.2 Deletion Syndrome. G3 (Bethesda). 2015 Nov;5(11):2453–61.

74. International Schizophrenia Consortium, Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009 Aug 6;460(7256):748–52.

Supplementary Material

Marker	rs1128349	rs11769825	rs3830638	rs2228607	rs2293484	rs6460052	rs86299	rs13246460	rs13232463	rs13241921	rs13246460	rs11770052	rs61742124	rs13227841	rs17851629	rs61438591	rs2240357	rs2074667
WBS1	C	G	G	A	G	A	C	A	NA	C	C	NA	C	C	G	C	C	A
WBS2	T	A	NA	G	NA	G	T	T	G	NA	T	T	T	T	G	C	C	A
WBS3	C	G	G	A	G	A	C	A	NA	C	C	A	C	C	A	T	T	G
WBS4	T	A	NA	G	C	G	T	T	G	T	T	T	T	T	A	T	T	G
WBS6	T	A	NA	G	C	G	T	T	G	T	T	T	T	T	G	C	C	A
WBS7	C	G	G	A	NA	A	C	A	C	NA	NA	NA	T	T	A	T	T	G
WBS8	C	G	G	NA	C	G	T	T	NA	T	T	NA	C	T	A	T	T	G
WBS5	C	A	G	A	G	A	C	A	C	C	C	A	C	C	A	T	T	G

Supplementary Table 1. Phased haplotypes of all individuals in the two linkage disequilibrium blocks identified from rs1138349 to rs13227841 (grey and blue) and from rs17851629 to rs2074667 (orange and green).

Individual	Gene	Chr	Start	End	Ref	Alt	AA Change	SIFT	Polyphen	KAVIAR_AF	pLI
WBS5	*PYHIN1*	1	158913694	158913694	C	T	NM_152501:p.R373X	NA	NA	2,8E-04	0,00
WBS6	*AGAP1*	2	236761328	236761328	A	T	NM_001244888:c.1051-2A>T	NA	NA	2,0E-04	1,00
WBS7	*USP45*	6	99936132	99936132	C	A	NM_001080481:p.E220X	NA	NA	6,5E-04	0,00
WBS8	*PIK3CG*	7	106508042	106508043	GA	0	NM_002649:p.12_13del	NA	NA	NA	0,99
WBS5	*UBR5*	8	103326142	103326142	G	A	NM_015902:p.R633X	NA	NA	NA	1,00
WBS4	*CC2D1A*	19	14031368	14031368	A	C	NM_017721:c.1357-2A>C	NA	NA	4,4E-04	0,00
WBS4	*DPYD*	1	97915674	97915674	T	G	NM_000110:p.K616Q	D	D	2,6E-05	0,00
WBS5	*SYN2*	3	12208742	12208742	C	T	UNKNOWN	D	D	6,5E-06	#N/A
WBS8	*ROBO2*	3	77638004	77638004	C	G	NM_002942:p.A868G	D	D		1,00
WBS7	*YEATS2*	3	183490204	183490204	G	T	NM_018023:p.V687F	D	D	6,5E-06	1,00
WBS3	*CTNND2*	5	10992686	10992686	C	T	NM_001332:p.R1063H	D	D	8,4E-05	1,00
WBS7	*PCDHAC2*	5	140347217	140347217	A	G	NM_018899:p.N289S	D	D	2,6E-05	0,85
WBS8	*HLA-A*	6	29910670	29910670	G	T	NM_001242758:p.E70D	D	D	NA	0,34
WBS1	*HLA-A*	6	29911901	29911901	C	A	NM_001242758:p.P208T	D	D	6,5E-06	0,34
WBS3	*ARID1B*	6	157488272	157488272	A	T	NM_017519:p.Q980L	D	D	3,8E-04	1,00
WBS3	*SDK1*	7	4272893	4272893	G	C	NM_001079653:p.G432A	D	D	5,5E-04	0,56
WBS1	*SDK1*	7	4285318	4285318	C	T	NM_001079653:p.R575W	D	D	5,2E-05	0,56
WBS8	*RELN*	7	103474008	103474008	G	A	NM_005045:p.A150V	D	D	6,5E-06	1,00
WBS7	*FOXP2*	7	114304331	114304331	G	A	NM_001172766:p.A614T	D	D	4,6E-04	1,00
WBS6	*PLXNA4*	7	131913203	131913203	G	A	NM_020911:p.R544W	D	D	NA	1,00
WBS1	*CHD7*	8	61778238	61778238	G	A	NM_017780:p.G2914R	D	D	3,0E-04	1,00
WBS2	*PAX5*	9	36966688	36966688	G	A	NM_016734:p.S213L	D	D	6,6E-04	0,98
WBS2	*KCNT1*	9	138641963	138641963	G	A	NM_020822:p.V92I	D	D	6,5E-06	0,82
WBS1	*PCDH15*	10	55571351	55571351	G	T	UNKNOWN	D	D	NA	0,00
WBS8	*PCDH15*	10	55780122	55780122	C	T	NM_001142765:p.V790M	D	D	6,2E-04	0,00
WBS8	*BDNF*	11	27680107	27680107	G	A	NM_170735:p.T2I	D	D	1,2E-03	0,94

Supplementary Table 2. Deleterious variants in ASD candidate genes per patient.

Individual	Gene	Chr	Start	End	Ref	Alt	AA Change	SIFT	Polyphen	KAVIAR_AF	pLI
WBS5	*KIRREL3*	11	126326293	126326293	C	T	NM_001161707:p.V267I	D	D	9,7E-05	0,95
WBS1	*CACNA1C*	12	2800220	2800220	A	G	NM_001129837:p.N2099S	D	D	5,2E-04	1,00
WBS3	*NBEA*	13	35731304	35731304	A	G	NM_015678:p.E914G	D	D	NA	1,00
WBS7	*DIAPH3*	13	60548565	60548565	T	C	NM_001258370:p.Q261R	D	D	1,2E-03	0,00
WBS3	*CHD8*	14	21897467	21897467	G	A	NM_001170629:p.L291F	D	D	8,9E-04	1,00
WBS6	*ESRRB*	14	76957891	76957891	G	A	NM_004452:p.D297N	D	D	2,4E-04	0,43
WBS7	*TSC2*	16	2108801	2108801	T	C	NM_000548:p.M301T	D	D	NA	1,00
WBS1	*TSC2*	16	2134981	2134981	C	T	NM_001077183:p.P1441L	D	D	NA	1,00
WBS5	*ZNF407*	18	72775366	72775366	G	A	NM_017757:p.G1897S	D	D	7,8E-05	1,00
WBS7	*ABCA7*	19	1051188	1051188	C	T	NM_019112:p.R907W	D	D	1,3E-05	0,00
WBS7	*TGM3*	20	2290392	2290392	C	T	NM_003245:p.R33W	D	D	NA	0,00
WBS4	*CBS*	21	44485800	44485800	C	A	NM_001178008:p.G153W	D	D	NA	0,02
WBS8	*CLTCL1*	22	19241639	19241639	A	G	NM_001835:p.L121S	D	D	1,3E-05	0,00

Supplementary Table 2 (continued)

.

Individual	Gene	Chr	Start	End	Ref	Alt	AA Change	SIFT	Polyphen	KAVIAR_AF	pLI
WBS8	*CELSR2*	1	109808789	109808789	G	A	NM_001408:p.G1992R	D	D	7,18E-04	1,00
WBS6	*MUC1*	1	155160698	155160698	A	T	NM_001018016:p.S66T	D	D	6,08E-04	0,91
WBS1	*ACTN2*	1	236902831	236902831	C	T	NM_001103:p.S369L	D	D	5,82E-05	1,00
WBS1	*SRSF4*	1	29475308	29475308	G	A	NM_005626:p.R367C	D	D	8,80E-04	0,97
WBS5	*PRDM16*	1	3328335	3328335	G	A	NM_022114:p.R525Q	D	D	1,16E-04	1,00
WBS7	*TAF5*	10	105127921	105127921	G	T	NM_006951:p.G59W	D	D	NA	1,00
WBS8	*BICC1*	10	60573675	60573675	G	A	NM_001080512:p.G821E	D	D	1,99E-03	0,90
WBS5	*SEC24C*	10	75529493	75529493	0	G	NM_004922:c.2682+2->G	NA	NA	NA	1,00
WBS5	*ZMIZ1*	10	81070780	81070780	A	T	NM_020338:p.S979C	D	D	4,72E-04	1,00
WBS8	*MUC5B*	11	1254407	1254407	G	A	NM_002458:p.G744S	D	D	1,18E-03	1,00
WBS5	*KIRREL3*	11	126326293	126326293	C	T	NM_001161707:p.V267I	D	D	9,70E-05	0,95
WBS8	*BDNF*	11	27680107	27680107	G	A	NM_170735:p.T2I	D	D	1,18E-03	0,94
WBS4	*CTNND1*	11	57561488	57561488	C	T	NM_001085462:p.R68W	D	D	4,07E-04	1,00
WBS7	*DPF2*	11	65113457	65113457	G	T	NM_006268:p.G278W	D	D	NA	1,00
WBS2	*RFX4*	12	107090184	107090184	G	T	NM_032491:p.G171W	D	D	NA	1,00
WBS3	*KSR2*	12	117914399	117914399	G	A	NM_173598:p.R789W	D	D	7,12E-05	1,00
WBS4	*NCOR2*	12	124821437	124821437	G	A	NM_001077261:p.R1983C	D	D	1,10E-04	1,00
WBS6	*NCOR2*	12	124841214	124841214	G	T	NM_001077261:p.P1062H	D	D	NA	1,00
WBS1	*CACNA1C*	12	2800220	2800220	A	G	NM_001129837:p.N2099S	D	D	5,24E-04	1,00
WBS3	*NBEA*	13	35731304	35731304	A	G	NM_015678:p.E914G	D	D	NA	1,00
WBS6	*PPP2R5C*	14	102229267	102229267	G	T	NM_001161725:p.G18V	D	D	NA	1,00
WBS7	*TECPR2*	14	102916924	102916924	G	A	NM_001172631:p.R1115H	D	D	NA	0,99
WBS2	*AKT1*	14	105241304	105241304	G	T	NM_001014431:p.L202M	D	D	NA	0,99
WBS3	*CHD8*	14	21897467	21897467	G	A	NM_001170629:p.L291F	D	D	8,93E-04	1,00
WBS7	*LEO1*	15	52245413	52245413	G	T	NM_138792:p.P463Q	D	D	NA	1,00
WBS4	*FAM214A*	15	52876954	52876954	A	G	NM_019600:p.L1022P	D	D	4,53E-05	1,00
WBS4	*CLPX*	15	65447268	65447268	A	G	NM_006660:p.I488T	D	D	1,33E-03	0,97
WBS6	*PML*	15	74336951	74336951	C	T	NM_033238:p.R751C	D	D	6,47E-06	0,97

Supplementary Table 3. Deleterious variants in functionally constrained genes with (pLI>0.9).

Individual	Gene	Chr	Start	End	Ref	Alt	AA Change	SIFT	Polyphen	KAVIAR_AF	pLI
WBS7	*TSC2*	16	2108801	2108801	T	C	NM_000548:p.M301T	D	D	NA	1,00
WBS1	*TSC2*	16	2134981	2134981	C	T	NM_001077183:p.P1441L	D	D	NA	1,00
WBS8	*ITFG1*	16	47294545	47294545	G	A	NM_030790:p.R378C	D	D	1,62E-04	0,98
WBS5	*CHD9*	16	53358047	53358047	C	G	NM_025134:p.A2629G	D	D	NA	1,00
WBS4	*KATNB1*	16	57787430	57787430	C	G	NM_005886:p.I392M	D	D	1,29E-05	1,00
WBS6	*VPS25*	17	40925830	40925830	A	C	NM_032353:p.K45Q	D	D	3,23E-05	0,91
WBS3	*MYH10*	17	8404273	8404273	T	A	NM_005964:p.Q1174H	D	D	NA	1,00
WBS7	*CXXC1*	18	47811697	47811697	0	GGAAG	NM_001101654:p.S222fs	NA	NA	NA	1,00
WBS5	*ZNF407*	18	72775366	72775366	G	A	NM_017757:p.G1897S	D	D	7,76E-05	1,00
WBS7	*KANK2*	19	11277247	11277247	G	A	NM_015493:p.S856L	D	D	6,66E-04	0,99
WBS8	*MED26*	19	16689144	16689144	A	G	NM_004831:c.147+2T>C	NA	NA	5,63E-04	0,93
WBS4	*JAK3*	19	17942530	17942530	G	A	NM_000215:p.R920C	D	D	NA	1,00
WBS4	*GRIN2D*	19	48902020	48902020	C	A	NM_000836:p.P124H	D	D	NA	1,00
WBS3	*PRR12*	19	50099036	50099036	C	0	NM_020719:p.P482fs	NA	NA	NA	1,00
WBS1	*PRPF31*	19	54627156	54627156	C	T	NM_015629:p.R186W	D	D	3,23E-05	0,95
WBS5	*SAFB*	19	5626443	5626443	G	C	NM_001201338:p.D73H	D	D	1,10E-04	1,00
WBS2	*MCM6*	2	136627912	136627912	G	A	NM_005915:p.R92W	D	D	9,12E-04	1,00
WBS4	*WIPF1*	2	175446073	175446073	G	A	NM_001077269:p.T49M	D	D	4,53E-05	0,98
WBS6	*AGAP1*	2	236761328	236761328	A	T	NM_001244888:c.1051-2A>T	NA	NA	2,01E-04	1,00
WBS3	*KIDINS220*	2	8871439	8871439	G	A	NM_020738:p.A1576V	D	D	9,70E-05	0,93
WBS4	*STK35*	20	2083502	2083502	C	A	NM_080836:p.P128H	D	D	NA	0,91
WBS5	*C20orf112*	20	31062448	31062448	T	A	NM_080616:p.D22V	D	D	1,29E-05	0,96
WBS8	*ZMYND8*	20	45839488	45839488	G	A	NM_012408:p.S1134L	D	D	2,65E-04	1,00
WBS7	*NCOA3*	20	46256365	46256365	G	A	NM_001174087:p.R198H	D	D	5,17E-05	1,00
WBS8	*NFATC2*	20	50133475	50133475	G	C	NM_001136021:p.L374V	D	D	NA	1,00
WBS1	*LAMA5*	20	60893962	60893962	G	A	NM_005560:p.R2327W	D	D	6,47E-05	1,00
WBS8	*LAMA5*	20	60921230	60921230	C	A	NM_005560:p.D442Y	D	D	NA	1,00
WBS5	*RTEL1*	20	62324279	62324279	C	A	NM_016434:p.S925Y	D	D	6,47E-06	1,00

Supplementary Table 3. (continued)

Individual	Gene	Chr	Start	End	Ref	Alt	AA Change	SIFT	Polyphen	KAVIAR_AF	pLI
WBS5	*USP25*	21	17135220	17135220	C	G	NM_001283041:p.T19R	D	D	2,98E-04	1,00
WBS8	*TIAM1*	21	32639180	32639180	G	A	NM_003253:p.R37W	D	D	1,94E-05	1,00
WBS3	*SGSM1*	22	25282688	25282688	G	C	NM_001098497:p.R588T	D	D	1,49E-04	0,92
WBS7	*CELSR1*	22	46835126	46835126	G	A	NM_014246:p.R1456W	D	D	2,01E-04	1,00
WBS8	*EPHB1*	3	134872994	134872994	C	0	NM_004441:p.A433fs	NA	NA	NA	1,00
WBS7	*EPHB1*	3	134977949	134977949	C	T	NM_004441:p.T981M	D	D	1,07E-03	1,00
WBS8	*FGD5*	3	14958775	14958775	A	G	NM_152536:p.Y1142C	D	D	1,10E-04	0,98
WBS7	*YEATS2*	3	183490204	183490204	G	T	NM_018023:p.V687F	D	D	6,47E-06	1,00
WBS6	*PAK2*	3	196529902	196529902	G	C	NM_002577:p.Q101H	D	D	6,47E-06	0,97
WBS1	*ARHGEF3*	3	56763366	56763366	C	T	NM_001128616:p.D511N	D	D	7,50E-04	0,98
WBS8	*ROBO2*	3	77638004	77638004	C	G	NM_002942:p.A868G	D	D	NA	1,00
WBS2	*ELMOD2*	4	141446611	141446611	A	G	NM_153702:p.Y10C	D	D	5,82E-05	0,93
WBS8	*RAPGEF2*	4	160277040	160277040	A	C	NM_014247:p.T1402P	D	D	NA	1,00
WBS8	*SLC4A4*	4	72338598	72338598	T	C	NM_003759:p.I561T	D	D	NA	1,00
WBS3	*CTNND2*	5	10992686	10992686	C	T	NM_001332:p.R1063H	D	D	8,41E-05	1,00
WBS2	*ADAMTS2*	5	178548667	178548667	G	C	NM_014244:p.S1058W	D	D	NA	0,97
WBS3	*SLC9A3*	5	476732	476732	G	A	NM_004174:p.R606W	D	D	9,12E-04	0,99
WBS3	*ARID1B*	6	157488272	157488272	A	T	NM_017519:p.Q980L	D	D	3,75E-04	1,00
WBS1	*BAG6*	6	31612929	31612929	G	T	NM_001098534:p.P388H	D	D	1,29E-05	1,00
WBS5	*COL11A2*	6	33133536	33133536	G	A	NM_080679:p.R1407C	D	D	6,47E-06	1,00
WBS6	*TFAP2D*	6	50712955	50712955	C	T	NM_172238:p.A340V	D	D	1,10E-04	0,97
WBS8	*RELN*	7	103474008	103474008	G	A	NM_005045:p.A150V	D	D	6,47E-06	1,00
WBS8	*PIK3CG*	7	106508042	106508043	GA	0	NM_002649:p.12_13del	NA	NA	NA	0,99
WBS7	*FOXP2*	7	114304331	114304331	G	A	NM_001172766:p.A614T	D	D	4,59E-04	1,00
WBS6	*PLXNA4*	7	131913203	131913203	G	A	NM_020911:p.R544W	D	D	NA	1,00
WBS2	*ABCB1*	7	87133705	87133705	G	A	NM_000927:p.R1233C	D	D	3,88E-05	0,93
WBS5	*UBR5*	8	103326142	103326142	G	A	NM_015902:p.R633X	NA	NA	NA	1,00
WBS5	*RIMS2*	8	105026804	105026804	C	A	NM_014677:p.Q913K	D	D	NA	1,00

Supplementary Table 3. (continued)

Individual	Gene	Chr	Start	End	Ref	Alt	AA Change	SIFT	Polyphen	KAVIAR_AF	pLI
WBS8	*ST18*	8	53044734	53044734	G	T	NM_014682:p.P817H	D	D	NA	1,00
WBS1	*CHD7*	8	61778238	61778238	G	A	NM_017780:p.G2914R	D	D	2,98E-04	1,00
WBS5	*SVEP1*	9	113173587	113173587	G	C	NM_153366:p.S2135C	D	D	1,77E-03	0,97
WBS2	*PAX5*	9	36966688	36966688	G	A	NM_016734:p.S213L	D	D	6,60E-04	0,98

Supplementary Table 3. (continued)

ID	Pvalue	Qvalue	ExAC frequency	WBS frequency	Gene	Overrepresented allele	eQTL loci	Individuals with risk allele
rs2135720	0,03	0,92	13395/66646	7/16	*PCDH15*	Risk		WBS2, WBS3 (hom), WBS5, WBS6, WBS7 (hom)
rs2642531	0,02	0,89	8349/58130	6/16	*MYBPH*	Risk		WBS1 (hom), WBS2, WBS4, WBS6, WBS7
rs2275477	0,01	0,78	14277/66662	8/16	*OSCP1*	Risk	*OSCP1*	WBS1, WBS3, WBS4 (hom), WBS5, WBS6, WBS7, WBS8
rs9960464	0,04	0,92	30575/66666	3/16	*UT2*	Protector	*SLC14A1*	WBS1, WBS5, WBS7
rs773852	0,04	0,92	26007/65326	2/16	*NWD1*	Protector		WBS7, WBS8
rs2042200	0,04	0,92	5653/66716	4/16	*ZNF426*	Risk	*ZNF426, ZNF121*	WBS3, WBS7 (hom), WBS8
rs4823086	0,05	0,92	8634/65914	5/16	*TBC1D10A*	Risk	*RP1.130H16.16*	WBS1, WBS4, WBS5, WBS6, WBS7
rs5749088	0,04	0,92	16432/66548	8/16	*RNF215*	Risk	*RP1.130H16.16*	WBS1, WBS3, WBS4 (hom), WBS5, WBS6, WBS7(hom)
rs2272739	0	0,67	402/66682	2/16	*LYAR*	Risk		WBS5, WBS6
rs396790	0,01	0,75	1773/66692	3/16	*FRAS1*	Risk		WBS1, WBS3, WBS7
rs450630	0,05	0,92	28943/66462	11/16	*SCAND3*	Risk		WBS1(hom), WBS2(hom), WBS3(hom), WBS5, WBS6(hom), WBS7, WBS8

Supplementary Table 4. Overrepresented previously associated alleles in our cohort.

ID	Pvalue	Qvalue	ExAC frequency	WBS frequency	Gene	Overrepresented allele	eQTL loci	Individuals with risk allele
rs2230365	0,01	0,78	11005/65348	7/16	*NFKBIL1*	Risk	*C4B*	WBS2, WBS3, WBS4, WBS6(hom), WBS7(hom)
rs2305598	0,03	0,92	31286/66716	3/16	*COL14A1*	Protector		WBS5, WBS6, WBS7, WBS8

Supplementary Table 4. (continued)

CHAPTER 5

Parental perspectives of knowledge, causes and recurrence risk in idiopathic autism and the effect of genetic counseling

Marta Codina-Solà, Miguel del Campo, Ivon Cuscó, Luis A. Pérez Jurado, Clara Serra-Juhé

In preparation

Abstract

Introduction: Autism Spectrum Disorders (ASD) are among the most inheritable neurodevelopmental disorders, but their aetiology remains unknown in the majority of cases. Therefore, most parents must base their reproductive decisions on empiric recurrence risk (RR) estimates. To determine factors influencing family planning and benefits of genetic counselling (GC), we studied knowledge and perceptions in parents with children with idiopathic ASD.

Materials and Methods: Parents with at least one child with ASD (n=39) answered a basal questionnaire addressing different topics, such as knowledge, perceived causes and RR or attitude towards genetics. A subset whose child had obtained negative results in molecular karyotype and exome sequencing (n=15) received GC and a following-up questionnaire afterwards and fifteen days later.

Results: Although most parents had received information about ASD, few had seen a medical geneticist or a genetic counsellor. Genetics was the most frequent perceived cause, especially among those with affected relatives. RR was overestimated by most parents and qualitative and quantitative estimates correlated. About half of parents believed that RR affected their family planning and their risk perception was higher compared to those who did not. After GC, quantitative but not qualitative RR estimates lowered and knowledge and favorable opinion towards genetics increased.

Conclusions: Although most parents perceived genetic factors as the most common cause of ASD, few had visited a genetics service. Perception of RR affects family planning and is overestimated by most parents. GC can improve knowledge and RR estimates and is key to take informed choices.

Introduction

Autism Spectrum Disorders (ASD) are a group of neurodevelopmental disorders characterized by impairments in three main domains: communication, social interaction and a restricted and repetitive pattern of behavior. They typically appear at early childhood and symptoms can vary greatly among affected persons, ranging from individuals with severe cognitive deficits to high functioning individuals with impaired language and poor social skills, but above average intelligence. They are estimated to affect 1 in 88 children in the United States, with a male to female ratio of 4:1(1). ASD are among the most inheritable neurodevelopmental disorders, with concordance rates of 60% between monozygotic twin pairs for classic autism, that increase to 90% when considering the Broader Autistic Phenotype (BAP) (2–9). They are considered a complex multifactorial disease, with great genetic heterogeneity and epigenetic and environmental factors contributing to the risk.

Current guidelines for genetic evaluation of patients with ASD recommend a tiered approach, in which pedigree analysis and physical examination to identify known syndromes should be the first steps. If the clinical presentation is suggestive of a specific syndrome, targeted testing of the responsible gene is recommended. If no particular condition is suspected, chromosomal microarray (CMA) is indicated for all patients. In addition, fragile X testing is recommended for all males, as well as *MECP2* sequencing to all females with a suggestive phenotype. Finally, *PTEN* testing should be performed in all patients with a head circumference >2.5 SD. Overall, the clinical genetics evaluation identifies an etiology in approximately 15-30% of individuals (10,11).

Therefore, most cases remain idiopathic and families must base their reproductive decisions on recurrence risk figures based on empirical estimates. When a specific pattern of inheritance cannot be identified, the empiric recurrence risk for ASD of unknown cause is thought to be 10% for classical ASD and 15-20% for milder associated conditions, usually termed the Broader Autistic Phenotype (BAP) (12–14). Families with more than one affected child, have a higher recurrence rate, approximately 35%. In addition, the degree of severity is unpredictable, whatever the diagnosis of the first child.

Genetic counseling, the process of helping people understand and adapt to the medical and familial implications of a genetic disease, might help families understand complex information. A correct understanding is key to adapt to a condition and to take informed choices, especially regarding family planning. Besides numerical recurrence risk figures, other factors that can affect decision-making in family planning are perceived recurrence risk, perceived causes and acceptance of genetic testing. Due to its complexity and the low diagnostic yield of current genetic tests, information regarding ASD heritability and recurrence risk can be difficult to understand and to assimilate. Therefore, genetic counseling has great relevance and is necessary both for families where a genetic cause can be identified and families where the cause remains idiopathic.

In this regard, several studies have shown that recurrence risk has a great impact on family planning in families with children with ASD of unknown cause. Reproductive stoppage, the trend for parents of a child with a severe disability to have none or fewer children, is a frequent phenomena among parents (13,15,16). In fact, studies have shown that reproductive behavior of affected families does not change for the first years after the first affected child is born, a time when the diagnosis has not been established, but after that, their birth rate falls compared to control families (16). Also, children with ASD are often the later to be born, suggesting that reproductive stoppage should be taken into consideration when estimating recurrence risk (13).

Since recurrence risk has a great influence on reproductive decisions, several studies have evaluated parents' perception, showing that it is consistently overestimated (17–20). Although the impact of perceived recurrence risk on reproductive behavior has been less explored, having a higher risk perception has been associated to a stronger influence on family planning (18). One of the factors influencing risk perception are perceived causes. Interestingly, genetic factors are among the most cited causes by parents, as well as, vaccines and pregnancy and delivery complications (17–19,21,22). Although studies in further populations are needed, cultural factors seem to have a great impact on perceived causes (20).

Despite the fact that ASD have a strong genetic component and a clinical genetic evaluation is recommended for all patients, several studies show an

underutilization of genetic services by affected families(18,23,24). Some studies have also addressed reasons to pursue further genetic testing among families with children with affected children(21). The majority of responders were favorable towards genetic testing, mainly to identify the cause of ASD, for family planning and to provide early intervention and treatment. The same study also suggested that parents who had undergone genetic testing often feel frustrated due to the low diagnostic yield of current diagnostic tests.

The present study extends previous research about knowledge and perceptions about ASD in a sample of Spanish parents, where similar studies had not been carried before. Particularly unique to this study is the study of the effect of genetic counseling in a group of parents who underwent CMA and exome sequencing in the context of a research project. Therefore, the aims of this study are: 1) To study perceived causes of ASD, knowledge of genetics, opinion and motivation for genetic testing in parents before genetic counseling sessions, 2) To assess recurrence risk perceptions and its impact on reproductive behavior and 3) To evaluate the impact of genetic counseling on knowledge, perceived personal control (PPC), perception of recurrence risk in a group of parents with children with idiopathic ASD.

Material and methods

Participants' recruitment

The study was aimed to parents with at least one child affected with ASD. Parents were recruited through three different settings. The first group consisted of parents who had attended the genetics or pediatric neurology services for the first time from January to May 2015 (n=9). Families with known diagnoses were excluded from the study and questionnaires were given to parents during waiting times before genetic counseling visits. The second group consisted of parents recruited through various local ASD family associations. Parents who agreed to participate sent their filled questionnaires back by email (n=15). Finally, a group of parents was recruited through a research study, which applied CMA and exome sequencing to unravel new genetic factors involved in the etiology of ASD. The group consisted of 15 parents from 11 families. Most families (n=9, 13 parents) had obtained negative results in both genetic tests.

To evaluate if genetic counseling affected parent's knowledge and perceptions of recurrence risk and causes of ASD, we carried out a pilot study in the third group of parents recruited through the research study (n=15). Parents, as well as close relatives to whom information may be relevant, were invited to attend a personalized genetic counseling session to discuss genetic tests performed during the project, results and implications for family members. Sessions included recording of medical and family history, discussion about genetic influence in ASD and modes of inheritance, genetic tests performed and implications for other family members. Sessions lasted from one hour and a half to two hours. A report explaining genetic tests, results and implications for other family members was sent to all participants (see Supplementary Information 1).

Instrumentation

The questionnaire was designed taking into account previous studies addressing parent's knowledge, beliefs and concerns regarding ASD and was reviewed by a genetic counselor and a clinical geneticist. Questions addressed different topics: 1) Demographics, 2) Previous information and sources of information, 3) Perceived causes of ASD, 4) Knowledge, opinions and motivation regarding genetics and genetic testing, 5) Knowledge and perception of recurrence risk and 6) Impact on reproductive behavior. Types of questions included different formats such as: dichotomous, multiple choice, Likert scales and nominal.

All parents were given a basal questionnaire addressing previously mentioned topics (Supplementary Information 2). Parents who attended genetic counseling sessions were also asked to answer the PPC questionnaire (25) before the genetic counseling session. After the genetic counseling session and two weeks later by phone interviews, they were asked to answer shorter version of the questionnaire addressing knowledge and perception of recurrence risk, applicable knowledge and opinion regarding genetics and PPC.

Data analysis

Descriptive statistics were calculated for most of the items. Correlations and statistical significance among selected items were assessed with chi-square and exact fisher's tests and t-tests for independent groups. Values in items regarding applied knowledge of genetics, opinion, perceptions of recurrence risk and PPC were compared before, right and fifteen days after genetic counseling sessions. Items of applied knowledge of genetics were rated on a one-point scale (0 points for incorrect or do not know answers and 1 point for correct answers). The total score was the sum of raw scores. Opinion's total score was calculated in a similar manner. Each item was rated on a one-point scale, and each item was rated on a three-point scale of agreement (-1=do not agree, 0= has not yet formed a view on this, 1=agree). Perceived personal control was measured as previously described (25) .

Results

Sample Demographics

A total of 39 parents agreed to participate in the study. As shown in table 1, the majority was female. Their mean age was 44.79 years (SD=7.35) and most (82%) was married. The highest level of education obtained varied among participants and most of them were employed full time. The majority of families had two children. All families had one affected child and most of them were reported to be affected with classic autism. Participants were also asked about the presence of additional affected family members and approximately 44% of respondents thought they had one or more affected relative.

Previous information about ASD

Most parents (29/37, 78%) had received previous information about ASD. When asked about which professional informed them about the disorder, a great majority answered that their source of information was a pediatrician neurologist (27/39, 69%). The second most frequent source of information was the Internet (9/39, 23%), followed by a clinical geneticist (4/39, 10%), a genetic counselor (2/39, 5%) and finally a primary care physician (1/39, 2.5%).

The most common topic parents had received information about was therapies and treatments (20/39, 51%). An important part had also received

information about diagnostic and characteristics of the disorder (16/39, 41%), but only a quarter or less had received information about recurrence risk (10/39, 25%), genetic tests available (8/39, 20%) and causes (4/39, 10%). Among parents who have been contacted through family associations, who represent an unbiased sample of the population, these percentages varied. Although the number of participants is small (n=15), none had seen a clinical geneticist.

Perceived causes of ASDs

To evaluate parents' perceived causes, participants were asked to choose as many items as wanted among a list of the most common reported causes of ASD. The most frequent believed cause was *genetic factors*, as 74% of parents (29/39) chose this option. The second most prevalent perceived cause was different *brain development* (15/39, 38%) followed by *vaccines* (10/39, 25%). A breakdown of the different causes included in the questionnaire and their relative percentage can be seen in Table 2

Gender	%
Men	26%
Women	74%
Age (years)	
26-34	28%
35-44	47%
45-49	15%
50-59	11%
Current relationship status	
Married	82%
Divorced	15%
Living with partner	3%
Highest level of education obtained	
Primary school	8%
High school	21%
Grade vocational training	28%
College graduate	36%
Post graduate degree	8%
Employment status	
Full time	49%
Half time	23%
Student	3%
Retired	0%
Work at home	18%
Unemployed	8%
Number of children	
1	21%
2	56%
3	23%
Diagnosis of affected children	
Autism	64%
Asperger	33%
PDD-NOS	3%

Table 1. Sample demographics

Cause	N	%
Genetic factors	29	74%
Alternative brain development	15	38%
Vaccines	10	26%
Pregnancy complications	9	23%
Toxic exposure during pregnancy	8	21%
Childbirth complications	3	8%
Childhood illness	2	5%
Infections during pregnancy	3	8%
Mother's age	1	3%

Table 2. Parents' perceptions of ASD causes.

Chi-square analysis was conducted to test whether perception of genetic factors as causative of ASDs was associated with reporting affected relatives and between reported diagnosis (classic autism, Asperger or PDD-NOS). Perception of genetic factors as a cause was associated with having an affected family member (p=0.017) and Asperger' self-reported status of affected child (p=0.01).

Knowledge and opinion about genetics

Next, we evaluated parents' knowledge and opinion about genetics. Questions were designed to evaluate topics that would have been addressed in a genetic counseling session and that could apply to most genetic diseases. Participants were asked to answer five true/false items, such as "Current genetic tests discard all causes of ASDs" or "Negative genetic test results mean that recurrence risk in an hypothetic subsequent pregnancy is similar to that of the general population". For 4 out of 5 of the questions, incorrect and do not know answers exceeded correct answers (Figure 1a).

Opinion about genetics was evaluated by four items addressing different topics, such as expectations, approval or interest in genetic testing. Participants were asked to rate each item with one of three possible options (Agree, disagree or do not know). Opinion about genetic testing was favorable, with positive answers towards genetic and genetic testing ranging from 70 to 80%. The lowest number of favorable opinions was registered for

an item regarding therapy development, which stated, “Genetic advances will contribute to development of therapies” (Figure 1b).

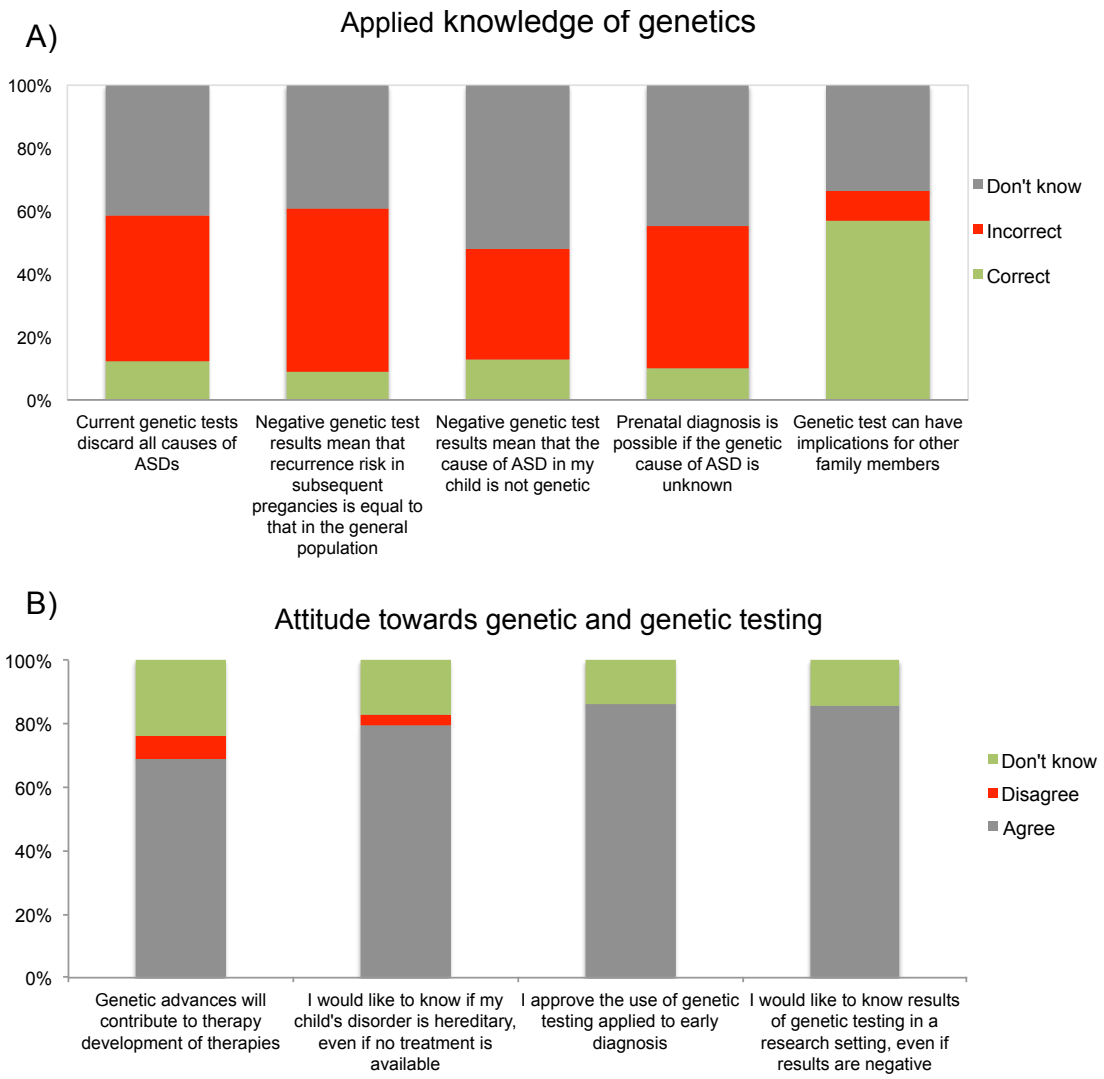


Figure 1. A) Applied knowledge of genetics. **B)** Opinion about genetics

Access and interest in genetic testing

Regarding access to genetic testing, 61% of participants (24/39) stated that their child had undergone some kind of genetic test. When asked what type of genetic test had their children undergone, 54% of participants (13/24) answered fragile X, followed by karyotype (9/24). Approximately 29% of respondents did not know what kind of genetic test had their children undergone. A similar proportion (29%, 7/24) answered MLPA and only 4% (1/24) answered CMA or single gene sequencing. Among parents recruited

through family associations, who are not biased by sample selection, only 27% (4/15) had access to genetic testing, specifically fragile X (n=3) and karyotype (n=2).

Interest in further genetic testing among parents with children with idiopathic ASD was high. The great majority (35/36) stated that they would be interested in further genetic testing. Parents were asked to indicate and prioritize reasons for wanting further testing. The participant who did not want further genetic testing stated that he/she "did not believed that the result would lead to a better treatment or therapy". Among parents who were interested in further genetic testing, the first reason was to improve medical follow-up of their affected children, followed by information for non-affected children (Table 3).

Reason for further genetic testing	1st choice (%)
Medical follow-up	31
Information for non affected children	21
Contribution to science	17
Definite diagnosis	10
Information for other family members	10
Future pregnancy	10

Table 3. First choice reason of parents for further genetic testing.

Perception of recurrence risk and effect on family planning

Perception of recurrence risk was measured in both quantitative and qualitative terms. Regarding quantitative risk, parents were asked to write their perceived recurrence risk from 0 to 100% in an open question. Qualitative risk perception was evaluated by a forced choice question, in which parents could choose between four different options: null, low, high or very high risk. Regarding quantitative measures, approximately a quarter of parents (26%) did not answer or said they did not know their recurrence risk (Table 4). Among parents who answered (n=29), 51% believed that their recurrence risk was 50% and a great majority (75%, 22/39) believed their recurrence risk was 50% or higher. Regarding qualitative measumrents, 54% of parents qualified their perceived recurrence risk as high (Table 5).

Quantitative risk range	N	%
0 -1%	1	3
1-10%	2	6
11-25%	4	10
26-49%	0	0
50-69%	15	38
70-99%	5	13
100%	2	5
N/K	3	8
N/A	7	18

Table 4. Quantitative risk estimates of recurrence risk among parents.

Qualitative risk	N	%
Very high	7	21
High	18	54
Low	6	18
Null	2	6

Table 5. Qualitative risk estimates of recurrence risk among parents.

Interestingly, perception of quantitative recurrence risk correlated with qualitative risk (Spearman R=0.75, p=0.01), although numeric values varied among each categorical risk classification (Figure 2).

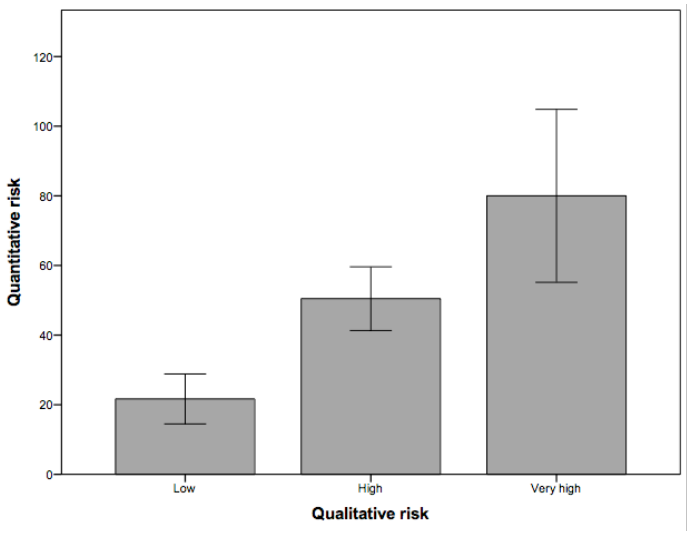


Figure 2. Graph showing mean quantitative risk within each category of qualitative risk. Error bars show 95% Confidence Intervals.

Regarding the effect of perceived recurrence risk on reproductive behavior, 54% of participants (20/37) answered that recurrence risk had had an effect on their family planning. Of these, 82% said that it had affected "much" their reproductive behavior. Statistical analyses were carried out to assess if higher risk perception was associated to self-reported effect on reproductive behavior. Student's t test was used to compare quantitative risk perception between respondents who answered that recurrence risk had an effect on their decisions and those who answered that it had not. Quantitative risk perception was not associated with family planning decisions, since mean quantitative recurrence risk measures did not differ between the two groups.

However, qualitative perception of recurrence risk was associated with reporting an effect on family planning. Fisher' exact test was used to test if the distribution of qualitative risk perception was significantly different between those who felt that it had an effect on family planning and those who did not. Parents who answered that perception of recurrence risk had an effect on family planning tended to perceive their risk as higher than those who did not (p=0.002, Figure 3).

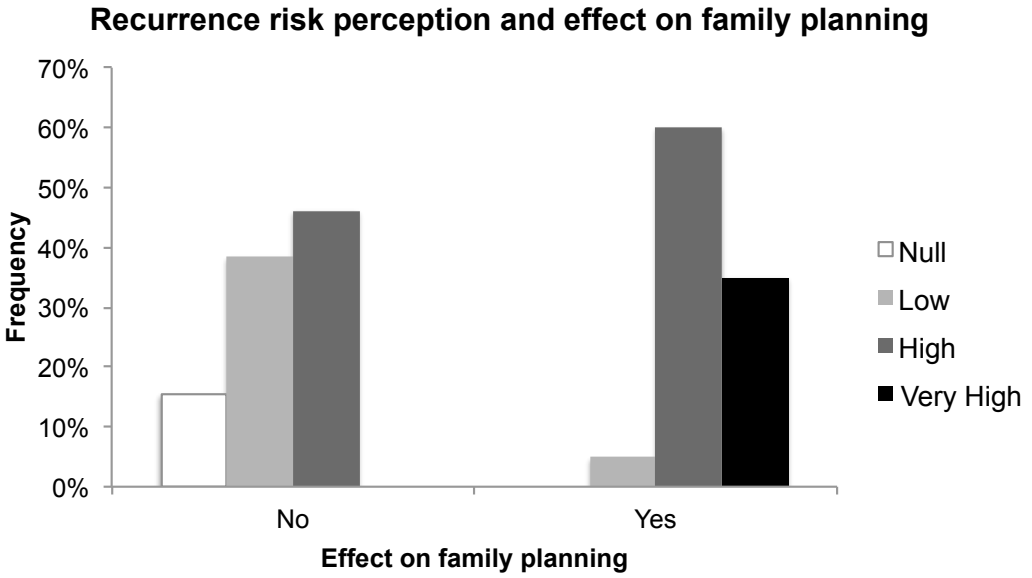


Figure 3. Distribution of qualitative risk among parents who stated that recurrence risk had had an effect on family planning and those who did not.

We also asked parents for reasons influencing their reproductive behavior (Table 6). The most common reason among respondents that believed recurrence risk had an effect on their family planning was “fear of having an affected children”, whereas among respondents not influenced by it (n=17), it was “family was already completed”.

Reason	Effect on family planning	
	Yes	No
Fear of having an affected child	17 (85%)	0
Economical resources devoted to a new child	8 (40%)	0
Time devoted to a new child	9 (45%)	1 (6%)
Efforts devoted to a new child	9 (45%)	0
Completed family	4 (20%)	12 (71%)

Table 6. Reasons influencing family planning among parents that believed that recurrence risk had an effect on their reproductive behavior and those who did not.

Effect of genetic counseling on knowledge, opinion, perceived recurrence risk and personal control

A subset of parents (n=15) from 11 families whose child had participated in a research project involving exome sequencing and CMA, were offered genetic counseling sessions to assess if genetic counseling improve knowledge about genetics, as well as about recurrence risk. Parents who agreed received genetic counseling sessions, which included recording of medical and family history, discussion about genetic influence in ASD and modes of inheritance, genetic tests performed and implications for other family members. Sessions lasted from one hour and a half to two hours. Participants received a shorter questionnaire before, right after the session and two weeks later to evaluate the effect of genetic counseling on items related to knowledge, perceived personal control (PPC), and recurrence risk perception.

Before genetic counseling sessions, the mean genetics knowledge score was 1.3/5 (SD=1.79), which increased to 3.9 after genetic counseling (SD=0.83) and to 3.64 (SD=1.11) two weeks after. The effect of genetic counseling in knowledge was statistically significant at both time points (Student’s t-test for paired samples, $p= 6x10^{-4}$ right after genetic counseling sessions and p=

$2x10^{-2}$ 15 days later, Figure 4A.) Regarding opinion about genetics and genetic testing, favorable attitude increase after the session, since the mean total score regarding opinion before genetic counseling was 2.47 (SD=1.88), and after was 3.5 (SD=0.63). However, this increase failed to reach statistical significance (p=0,051, Figure 4B).

Mean perceived quantitative recurrence risk was compared before and after genetic counseling sessions. Families who were given positive results (n=2) were excluded, since their recurrence risk varied from those with children with idiopathic ASD. Since the cause was *de novo* in both families, their risk was estimated at 1% taking into account the possibility of germinal mosaicism. Participants from both families (n=2) estimated their recurrence risk correctly after and 15 days after genetic counseling session. Regarding the rest of participants with negative results, all estimated their recurrence risk at 50% before genetic counseling sessions. After genetic counseling sessions mean perceived recurrence risk was reduced to 24.5% (SD=15%, (p=0.001), Figure 4C). Moreover, the great majority of participants (77%, 10/13) estimated that their risk was lower compared to their initial belief and of these, 7/10 (70%) gave accurate numeric values. However, 15 days after genetic counseling sessions, perceived risk increased for 5/10 (50%) of the participants who had lowered their estimates after sessions. Mean perceived risk 15 days after sessions was 30.8% (SD=28%). This was not statistically significant (p=0.06), although it shows a certain tendency.

Qualitative measures of perceived recurrence risk were compared before, after genetic counseling session and 15 days later using the Wilcoxon rank test. Again, participants with positive results were excluded from the analysis. Both participants qualified their risk as null, since the cause was found to be *de novo* in both cases. Regarding cases with negative results, qualitative measures before genetic counseling sessions did not differ significantly from those after sessions (Wilcoxon signed-rank test, p= 0.56) or 15 days later (Wilcoxon signed-rank test, p= 0.26). However, post counseling qualitative high and low values seem to increase both just and 15 days the after sessions (Figure 4D).

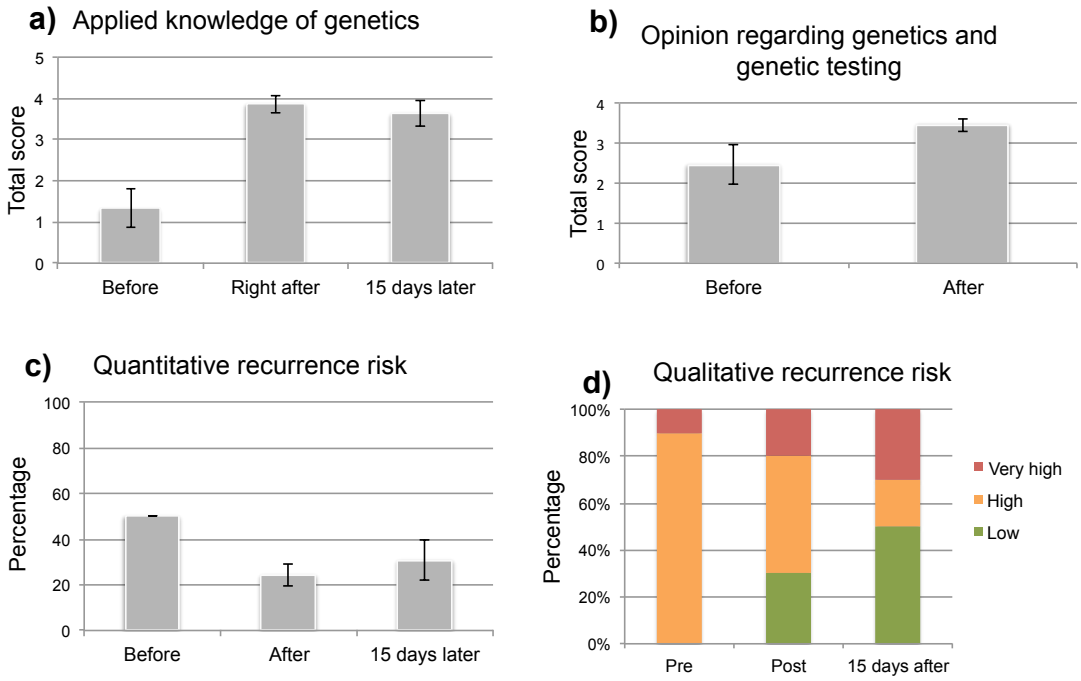


Figure 4 a) Total knowledge scores pre, post and 15 days after genetic counseling sessions. **b)** Total opinion scores pre, post and 15 days after genetic counseling sessions. **c)** Quantitative recurrence risk pre, post and 15 days after genetic counseling sessions. **d)** Qualitative perception of recurrence risk pre, post and 15 days after genetic counseling sessions.

Since PPC scores have been reported to increase after definite diagnosis, parents with positive results (n=2) were excluded from this analysis. PPC total score mean before genetic counseling was 1.30/2 (SD=0.59) and this score was slightly increased after genetic counseling (mean= 1.37, SD=0.56). This result was not statistically significant (p=0.093), although it approaches significance. Comparisons of cognitive, decisional and behavioral PPC dimensions were neither statistically significant, although the decisional mean score increased slightly, but failed to reach significance (p=0.096).

Discussion

In this study, we evaluated previous information, knowledge, opinion, perceived causes and recurrence risk in a group of 39 parents with a child affected with ASD. In addition, we assessed the effect of genetic counseling in a group of participants whose child obtained negative results in two of the current tests of choice for the genetic diagnosis of ASD (CMA and exome sequencing). Overall, the results show an underutilization of genetic services,

although nearly all participants expressed their interest in it and genetics was the most common cited cause by parents. Moreover, most parents had an inaccurate perception of their recurrent risk, yet half of them admitted it had an effect on their reproductive decisions. Finally, we show that genetic counseling has a positive effect on recurrence risk estimates, both for families with definite genetic diagnosis and those with ASD of unknown cause.

Similar to previous results, we report an underutilization of genetic services by families with children with ASD (17,18,24). Only 10% of parents in our study had seen a clinical geneticist and, none of the participants recruited through families associations had access to a genetics service. It is also of interest, that two of the participants said they had never received previous information about ASD, despite having visited several health professionals multiple times. This suggests that the information provided during those visits did not meet parents' needs and so, was perceived as of no utility.

Clearly, the lack of utilization of genetic services does not seem to be due to a disinterest, since the great majority of parents in our study (74%) believed that genetic factors were one of the main causes of ASDs, similar to previous studies conducted in Western countries (17,18). Perception of genetic factors as a cause of ASD was related to presence of affected family members, highlighting that personal experience is one of the main factors that shape perception and beliefs. Interestingly, that 26% of parents in our study believed vaccination was a causative factor for ASD, similar to previous reports (17,18). Since vaccines are key elements of public health, health professionals should address cause perceptions during visits and provide parents with updated information about its benefits and the lack of proven association.

Despite the fact that very few parents reported seeing a genetic professional, 60% of families had undergone genetic testing. The percentage of families who had received genetic testing was lower among parents recruited through family associations, who represent an unbiased sample, since only 26% had received some kind of genetic testing. Among this unbiased sample, the only genetic testing received was karyotype and fragile X testing and none of the participants had received CMA, the recommended first-choice test in ASD.

Interestingly, most participants showed a lack of knowledge about basic genetic concepts. Parents were asked several questions about the meaning of negative results and its implications for recurrence risk and reproductive options, showing that for most of the answers, incorrect and do not know answers outnumbered correct answers. This suggests that these topics were not addressed during pretest counseling sessions.

Despite this lack of access to genetic services, interest in genetic testing was very high. Specifically, all participants except one expressed their interest in further genetic testing. The most common reason was to improve the medical management of their affected child and only 10% of parents would like to use this information for family planning. This low percentage may be due to time elapsed between ASD diagnosis and the present study and the age of our participants, since 60% of the participants were older than 40 years old. Moreover, the fact that half of the participants admitted that recurrence risk had an effect on family planning, suggests that genetic testing and conseling would have been considered useful if offered at the right time. In agreement, among families who felt that recurrence risk had had an effect on family planning, the main reason influencing reproductive behavior was “fear of having an affected child”. On the contrary, among parents who stated that it had not an effect it was “family already completed”. The high influence of recurrence risk on reproduction agrees with previous studies reporting a high incidence of reproductive stoppage in families with children with ASDs (13,15,16).

The high influence of recurrence risk on family planning is also related to the inaccurate overestimation of recurrence risk by parents, which has also been reported in previous studies (18,20). In our study, most parents estimated their recurrence risk in 50%, a calculation based on the relation between the number of affected children and the total number of children they had. Interestingly, quantitative risk correlated with qualitative risk, with higher numeric estimates being more frequent among categorical estimates of “high” or “very high” risk. Therefore, perceived recurrence risk seems to be one of the most important decision making factors among parents with a child on the spectrum.

As genetic counseling might help families understand complex information and adapt to this risk, we explored if genetic counseling had an effect on knowledge, recurrence risk estimates and PPC. To this end, we offered counseling sessions to a group of parents (n=15) who had undergone exome sequencing and CMA in a research study. Despite our small sample size, the results show that genetic counseling improved knowledge scores both after sessions and 15 days after. PPC increased slightly after sessions, although the difference was not statistically significant. This may be due the fact that all parents had obtained negative results in both tests, since previous studies have shown that higher increments in PPC are experienced when families are given positive results. Although PPC did not change after genetic counseling sessions, recurrence risk perception lowered after sessions and two weeks later. Interestingly, qualitative risk estimates decreased from high to low for a group of respondents, but increased from high to very high for another group two weeks after genetic counseling sessions. This may be explained by different personal experiences between the two groups that shape their recurrence risk perceptions. Also, the qualitative difference between the categories “high” and “very high” may seem smaller than between “high” and “low”, making it more probable for parents to shift between the first two categories than between the last. Finally, the time elapsed between genetic counseling sessions and phone interviews two weeks after, may allow parents to assimilate the information provided and change their recurrence risk perception. In addition, sessions also provided parents with a space to share their experiences, which was corresponded with a general feeling of gratitude. Most participants recalled their initial confusion after the diagnosis and stated that they would have welcomed professional guidance, as well as an integral care plan. Siblings who attended counseling sessions shared their worries about the responsibility that entailed taking care of their affected sibling, as well as fears for their own recurrence risk.

To conclude, the results of this study show an underutilization of genetic services among parents with children with ASD. In contrast, most parents believe genetics is one of the major causes and express interest in genetic testing. Despite this, most families had not access to CMA, the standard and recommended test of choice. The underutilization of genetic services leads,

among others, to an overestimation of recurrence risk, which has a great impact on family planning. Our smaller study on a selected group of parents shows that genetic counseling can help parents to understand the implications of negative results, as well as to better estimate their recurrence risk. Currently, about 20% of children with ASDs can be diagnosed by a tiered diagnostic approach, and this percentage will surely increase in the next years with the advent of next generation sequencing. Therefore, an important number of families would benefit from a comprehensive genetic evaluation and genetic counseling. For all that, the referral criteria and diagnostic process of children with ASD should be revised, to ensure that all families have access to genetic services and genetic counseling.

References

1. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators, Centers for Disease Control and Prevention, Baio J et al. Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. MMWR Surveill Summ. 2012 Mar 30;61(3):1–19.
2. Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. J Child Psychol Psychiatry. 1977 Sep;18(4):297–321.
3. Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM. Concordance for the syndrome of autism in 40 pairs of afflicted twins. Am J Psychiatry. 1985 Jan;142(1):74–7.
4. Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, et al. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. J Child Psychol Psychiatry. 1989 May;30(3):405–16.
5. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, et al. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med. 1995 Jan;25(1):63–77.
6. Taniai H, Nishiyama T, Miyachi T, Imaeda M, Sumi S. Genetic influences on the broad spectrum of autism: study of proband-ascertained twins. Am J Med Genet B Neuropsychiatr Genet. 2008 Sep 5;147B(6):844–9.
7. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. Arch Pediatr Adolesc Med. 2009 Oct;163(10):907–14.
8. Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry. 2010 Nov;167(11):1357–63.
9. Ronald A, Hoekstra R a. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156(3):255–74.
10. Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. Genet Med. 2013 May;15(5):399–407.

11. Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, et al. Clinical genetic testing for patients with autism spectrum disorders. Pediatrics. 2010;125(4):e727–35.
12. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics. 2011 Sep;128(3):e488–95.
13. Wood CL, Warnell F, Johnson M, Hames A, Pearce MS, McConachie H, et al. Evidence for ASD recurrence rates and reproductive stoppage from large UK ASD research family databases. Autism Res. 2015 Feb 1;8(1):73–81.
14. Miles JH, McCathren RB, Stichter J, Shinawi M. Autism Spectrum Disorders. University of Washington, Seattle; 2010.
15. Grønborg TK, Hansen SN, Nielsen S V, Skytthe A, Parner ET. Stoppage in Autism Spectrum Disorders. J Autism Dev Disord. 2015 Nov;45(11):3509–19.
16. Hoffmann TJ, Windham GC, Anderson M, Croen LA, Grether JK, Risch N. Evidence of reproductive stoppage in families with autism spectrum disorder: a large, population-based cohort study. JAMA psychiatry. 2014 Aug;71(8):943–51.
17. Mercer L, Creighton S, Holden JJ a, Lewis MES. Parental perspectives on the causes of an autism spectrum disorder in their children. J Genet Couns. 2006 Feb;15(1):41–50.
18. Selkirk CG, McCarthy Veach P, Lian F, Schimmenti L, LeRoy BS. Parents' perceptions of autism spectrum disorder etiology and recurrence risk and effects of their perceptions on family planning: Recommendations for genetic counselors. J Genet Couns. 2009 Oct;18(5):507–19.
19. Whitelaw C, Flett P, Amor DJ. Recurrence risk in Autism Spectrum Disorder: a study of parental knowledge. J Paediatr Child Health. 2007 Nov;43(11):752–4.
20. Chen LS, Li C, Wang CH, Amuta a, Li M, Huang TY, et al. Autism spectrum disorders: perceptions of genetic etiology and recurrence risk among Taiwanese parents of affected children. Clin Genet. 2014 Sep 30;1–6.
21. Chen L-S, Xu L, Huang T-Y, Dhar SU. Autism genetic testing: a qualitative study of awareness, attitudes, and experiences among parents of children with autism spectrum disorders. Genet Med. 2013 Apr;15(4):274–81.
22. Harrington JW, Patrick P a, Edwards KS, Brand D a. Parental beliefs about autism: implications for the treating physician. Autism. 2006 Sep;10(5):452–62.
23. Shea L, Newschaffer CJ, Xie M, Myers SM, Mandell DS. Genetic testing and genetic counseling among Medicaid-enrolled children with autism spectrum disorder in 2001 and 2007. Hum Genet. 2014 Jan;133(1):111–6.
24. Vande Wydeven K, Kwan A, Hardan AY, Bernstein J a. Underutilization of genetics services for autism: the importance of parental awareness and provider recommendation. J Genet Couns. 2012 Dec;21(6):803–13.
25. Berkenstadt M, Shiloh S, Barkai G, Katznelson MBM, Goldman B. Perceived personal control (PPC): A new concept in measuring outcome of genetic counseling. Am J Med Genet. 1999;82(1):53–9.

Supplementary Information 1

Máster en Asesoramiento Genético
Universitat Pompeu Fabra

CUESTIONARIO DEL PROYECTO: "ASESORAMIENTO GENÉTICO EN TRASTORNOS DEL ESPECTRO AUTISTA"

Este cuestionario tiene como objetivo el análisis del conocimiento, opiniones y necesidades de las familias con un hijo con trastorno del espectro autista (TEA) en relación al asesoramiento genético. Los datos obtenidos mediante este estudio serán tratados de forma completamente anónima. Este estudio tiene como objetivo determinar cuales son las necesidades de las familias con el fin de mejorar la calidad del asesoramiento genético en familias con hijos con trastorno del espectro autista.

Por favor, conteste siguientes preguntas marcando con una X las casillas correspondientes o rellenando las tablas correspondientes. Cuando termine, haga clic en guardar.

1. ¿Ha recibido información anterior sobre los trastornos del espectro autista?

- ☐ Sí
- ☐ No

En caso afirmativo, ¿qué profesional le ha transmitido y asesorado sobre este tema? **Puede marcar más de una casilla.**

- ☐ Mi médico de cabecera
- ☐ Un médico especialista: pediatra, neurólogo, etc.
- ☐ Un genetista clínico
- ☐ Un asesor genético
- ☐ Un enfermero/a
- ☐ A través e internet o otros medios de comunicación
- ☐ Otros (**por favor, especifique**):

En caso afirmativo, ¿qué información ha recibido a través de este/os profesionales? **Puede marcar más de una casilla.**

- ☐ Información sobre el riesgo de recurrencia para un futuro embarazo
- ☐ Información sobre las causas de la enfermedad
- ☐ Información sobre el diagnóstico y la historia natural de la enfermedad
- ☐ Información sobre terapias y tratamientos para la enfermedad
- ☐ Información sobre las pruebas genéticas disponibles para su diagnóstico

2. ¿Tiene contacto o forma parte de alguna asociación de padres con hijos/as con TEA o persones con TEA?

- ☐ Sí
- ☐ No

En caso afirmativo, ¿de qué asociación se trata?

3. ¿Cuáles de estos factores cree usted que pueden causar autismo?
Puede marcar más de una casilla.

- ☐ Complicaciones durante el parto
- ☐ Complicaciones durante el embarazo
- ☐ Dieta (alergias o intolerancias a ciertos alimentos, etc.)
- ☐ Exposición a tóxicos durante el embarazo (tabaco, drogas, alcohol, radiación…)
- ☐ La edad del padre en el momento del nacimiento
- ☐ Factores genéticos
- ☐ Enfermedades durante la infancia o la niñez
- ☐ Diferente desarrollo cerebral
- ☐ Infecciones durante el embarazo
- ☐ La edad de la madre en el momento del nacimiento
- ☐ Vacunas
- ☐ Otros (por favor, especifique):

4. ¿Sabe si a su hijo se le ha realizado algún tipo de estudio genético?

- ☐ Sí
- ☐ No
- ☐ No lo sé

En caso **afirmativo**, ¿qué pruebas genéticas le han sido realizadas?

- ☐ Cariotipo
- ☐ aCGH o array
- ☐ X-frágil
- ☐ MLPA
- ☐ Secuenciación de exoma
- ☐ Secuenciación de genes específicos
- ☐ No lo sé

5. ¿En relación a los estudios genéticos realizados o no en su hijo/a, cuáles de estas afirmaciones le parecen verdaderas o falsas? Marque con una X la casilla correspondiente.

	Verdadero	**Falso**	**No lo sé**
a. Las pruebas genéticas realizadas en mi hijo/a descartan todas las posibles causas genéticas.			
b. El hecho que todos las pruebas genéticas hayan dado negativo en mi hijo/a significa que el riesgo de recurrencia para un siguiente embarazo es muy bajo, igual al de otra pareja sin un antecedente.			
c. El hecho que todos las pruebas genéticas hayan dado negativo en mi hijo/a significa que su autismo no tiene un origen genético			
d. Se puede realizar diagnóstico prenatal en un siguiente embarazo aunque no se conozca la anomalía genética responsable de la enfermedad en el primer hijo afecto			
e. El resultado de un test genético puede tener implicaciones para otros familiares			

6. ¿En relación a los estudios genéticos realizados en su hijo/a, con cuáles de estas afirmaciones está de acuerdo? Marque con una X la casilla correspondiente.

	De acuerdo	**En desacuerdo**	**No lo sé**
a. Creo que el desarrollo de la genética contribuirá al desarrollo de tratamientos para el trastorno del espectro autista.			
b. Me gustaría saber si la enfermedad de mi hijo es hereditaria, aunque no haya tratamiento disponible			
c. Apruebo el uso de pruebas genéticas que puedan contribuir al diagnóstico temprano de la enfermedad			
d. En el caso de participar en estudios de investigación adicionales, me gustaría ser informado de los resultados, aunque no se haya encontrado ningún resultado			

7. Si fuera posible, ¿le gustaría participar en estudios genéticos adicionales que permitieran determinar la causa del autismo en su hijo/a?

☐ Sí
☐ No

En caso **afirmativo**, ¿cuáles serían sus principales motivaciones?
Ordénelas de mayor a menor preferencia, donde 1 es el principal motivo y 7 es el motivo que menos le influiría:

Número	**Motivación**
	Contribuir al conocimiento y a la ciencia
	Obtener un diagnóstico definitivo
	Beneficios para otros familiares: establecer su riesgo de tener un hijo con autismo
	Establecer el riesgo de recurrencia para el siguiente embarazo con el fin de tener un hijo no afecto
	Beneficios para los hijos no afectos: establecer su riesgo de tener un hijo con autismo
	Mejorar el manejo y tratamiento de mi hijo con autismo
	Otros (**por favor especifique**):

En caso **negativo**, ¿cuáles serían sus principales motivaciones?

☐ No creo que el resultado del test mejore la situación de mi hijo ni contribuya a un mejor tratamiento
☐ No deseo saber si se trata de una enfermedad hereditaria
☐ No creo que la causa del autismo de mi hijo/a sea de origen genético
☐ Otros (por favor especifique):

8. Si tuviera otro hijo, escriba el riesgo del 0% al 100% que cree que tendría padecer autismo, donde 0% es nulo y 100% es máximo:

[ ] %

9. ¿Si tuviera otro hijo/a, cuál cree que sería su riesgo de padecer autismo?

- [ ] Riesgo cero
- [ ] Riesgo bajo
- [ ] Riesgo alto
- [ ] Riesgo muy alto

10. En caso de tener otro hijo y resultar afecto, ¿cree usted que su autismo sería…

- [ ] de menor gravedad
- [ ] de igual gravedad
- [ ] de mayor gravedad
- [ ] No es posible prever la gravedad
- [ ] No lo sé

11. ¿Cree usted que el riesgo de tener otro hijo ha afectado a su decisión de tener más hijos?

- [ ] Sí
- [ ] No

En caso afirmativo, ¿en qué grado cree usted que afecta a su decisión?

- [ ] Poco
- [ ] Bastante
- [ ] Mucho

12. ¿Qué motivos influyen o influyeron en su decisión de tener o no tener más hijos/as?
Puede marcar más de una casilla.

- [ ] El miedo de tener otro hijo/a afecto
- [ ] Tener otro hijo/a me impediría dedicar el tiempo necesario para el cuidado de mi hijo con autismo
- [ ] Tener otro hijo/a me impediría dedicar el dinero necesario para el cuidado de mi hijo con autismo
- [ ] Tener otro hijo/a me impediría dedicar el esfuerzo necesario para el cuidado de mi hijo con autismo
- [ ] No deseo tener más hijos/as independientemente del resultado
- [ ] Otros (por favor, especificar):

DATOS DEMOGRÁFICOS

1. Edad actual:

2. Relación actual
- ☐ Divorciados o separados
- ☐ Pareja de hecho
- ☐ Casados
- ☐ Soltero/a
- ☐ Víduo/a
- ☐ Otros (especificar): _____

3. Grado máximo de educación:

- ☐ Primaria
- ☐ Instituto
- ☐ Formación profesional
- ☐ Diplomatura/licenciatura
- ☐ Postgraduado (máster, postgraduado)
- ☐ Educación postgraduado (doctorado)

4. ¿Cuál es su estado laboral actualmente?

- ☐ Trabajo a tiempo completo
- ☐ Trabajo a media jornada
- ☐ Estudiante
- ☐ Jubilado
- ☐ Trabajo en el hogar
- ☐ Sin trabajo
- ☐ Otros (especificar):

6. Por favor, rellene la siguiente tabla respecto al diagnóstico y edad de sus hijos:

Hijo/a	**Edad**	**Sexo (H/M)**	**Diagnóstico: afecto de autismo (SÍ/NO)**
1			
2			
3			
4			
5			

7. ¿Alguno de sus otros familiares padece también autismo?

- ☐ Sí
- ☐ No
- ☐ No lo sé

En caso afirmativo, ¿qué relación familiar tiene con él o ella?

Si desea recibir más información e asesoramiento genético acerca de los Trastorno del Espectro Autista, por favor deje sus datos de contacto:

Nombre y apellidos:
Teléfono/s de contacto:
Dirección de correo electrónico (opcional):

Supplementary Information 2

INFORME DE ASESORAMIENTO GENÉTICO

En la Unidad de Genética de la Universitat Pompeu Fabra se ha llevado a cabo un estudio de investigación con el objetivo de determinar las bases genéticas de los trastornos del espectro autista (TEA). Este estudio ha sido financiado por agencias públicas (FIS/ISCIII) y autorizado por el Comité Ético de Investigación Clínica de la Institución. Las muestras se obtuvieron con consentimiento informado. Se ha recibido una muestra de sangre periférica del paciente PACIENTE , a partir de la cual se ha extraído ADN y se han realizado dos estudios genéticos (cariotipo molecular y secuenciación de exoma) con finalidades diagnósticas y de investigación.

Nombre del paciente: PACIENTE
Hospital de referencia: Hospital del Mar

MOTIVO DE CONSULTA

Acuden a consulta PADRE y MADRE, padres de PACIENTE, con el fin de conocer los resultados de las pruebas genéticas de PACIENTE.

ANTECEDENTES PERSONALES

PACIENTE es fruto de una gestación controlada, sin incidencias. Su desarrollo psicomotor fue normal durante el primer año. A los dos años, fue derivado al CDIAP por un retraso en el desarrollo psicomotor, donde estuvo en seguimiento hasta los cinco años. De pequeño realizaba con frecuencia estereotipias, que han disminuido mucho con la edad. En el año 2002 fue diagnosticado de TEA clásico mediante la Entrevista para el Diagnóstico de Autismo-Revisada (ADI-R), cumpliendo puntuaciones en las tres áreas evaluadas (comunicación, interacciones sociales recíprocas y conductas e intereses restringidos, repetitivos y estereotipados). Actualmente, PACIENTE tiene 10 años. En cuanto al lenguaje, presenta una buena comprensión y un lenguaje expresivo más limitado.

Pruebas genéticas anteriores:

- Cariotipo: 46, XY.
  El cariotipo permite estudiar alteraciones del número de cromosomas, así como pérdidas, ganancias y reordenamientos del material genético de gran tamaño. Los cromosomas son las estructuras en las cuales se organiza nuestro material genético dentro de las células. Cada una de nuestras células contiene 46

cromosomas (23 pares), 23 procedentes de la madre y 23 procedentes del padre. En cada cromosoma hay varios genes o unidades de información y de cada uno de ellos también tenemos dos copias, la que heredamos del padre y la que heredamos de la madre. En este caso, el estudio del cariotipo no reveló ninguna alteración numérica o estructural.

- Estudio molecular de X-frágil (Hospital de Cruces): negativo.
  El síndrome de X-frágil es la causa más frecuente de discapacidad intelectual hereditaria y una de las causas más comunes de TEA. Se estima que entre el 2-6% de los casos de TEA son debidos al síndrome de X-frágil y que un 30% de los afectos de X-frágil tienen TEA. El estudio molecular del síndrome de X-frágil no reveló ninguna alteración en el gen *FMR1,* descartando este síndrome como causante del trastorno de PACIENTE.

ANTECEDENTES FAMILIARES:

PACIENTE es hijo de una pareja sana no consanguínea. Tiene un hermano mayor y una hermana menor, HERMANO_VARÓN y HERMANA_MUJER. No hay antecedentes familiares de interés.

PRUEBAS REALIZADAS EN LA UNIDAD:

- Cariotipo molecular (diseño propio, 5300 BACs, densidad cada 0.5Mb): negativo
  El cariotipo molecular detecta pérdidas o ganancias de material genético de menor tamaño no visibles mediante cariotipo convencional. Actualmente es la prueba genética recomendada para el diagnóstico genético de trastornos del espectro autista. En el caso de PACIENTE, el estudio del cariotipo molecular no ha revelado ninguna ganancia ni pérdida de material genético responsable del trastorno.

- Secuenciación de exoma (captura Nimblegen SeqCap EZ Exome v2.0-36 Mb, secuenciador SOLiD 4 Seq, Applied Biosystems): negativo
  La secuenciación de exoma es una técnica que permite estudiar todos los genes en un solo análisis. Detecta pequeños cambios del material genético, llamadas variantes o mutaciones, que pueden alterar la función de un gen y dar lugar a enfermedad. Este estudio no ha revelado la presencia de ninguna mutación asociada con el

trastorno de PACIENTE y no ha permitido establecer si existe una alteración genética responsable del fenotipo.

VALORACIÓN Y ASESORAMIENTO GENÉTICO

Los trastornos del espectro autista o TEA son un grupo de trastornos del neurodesarrollo infantil que normalmente se manifiestan antes de los 3 años de edad. Aunque los síntomas del TEA varían mucho en cada niño o niña, se pueden agrupar en tres áreas del desarrollo: conductas repetitivas y estereotipadas, dificultades de comunicación y dificultades en las relaciones sociales. Estas alteraciones se expresan de forma distinta en cada niño y, por eso, se habla de un grupo o espectro de trastornos. Tienen una prevalencia estimada de 1 de cada 110 niños y son cuatro veces más frecuentes en varones que en mujeres. Algunos niños y niñas con TEA pueden tener también otras alteraciones cognitivas, como dificultades de aprendizaje o discapacidad intelectual, trastorno por hiperactividad y déficit de atención o epilepsia.

Todavía no se conocen todas las causas TEA, pero se cree que están involucrados tanto factores genéticos como ambientales. El material genético es el libro de instrucciones de nuestro cuerpo y regula tanto su formación como su funcionamiento. El material genético se encuentra en cada una de nuestras células y se organiza en 46 cromosomas (23 pares), 23 procedentes de la madre y 23 procedentes del padre. En cada cromosoma hay varios genes o unidades de información y de cada uno de ellos también tenemos dos copias, la que heredamos del padre y la que heredamos de la madre. Desde el punto de vista genético, el TEA es una entidad muy heterogénea, en la que pueden estar implicados cambios genéticos o mutaciones en centenares de genes. Actualmente, se considera que la gran mayoría de casos (alrededor del 70%) son debidos a un **modelo multifactorial**. En los casos multifactoriales, la suma de cambios en el material genético de pequeño efecto y de factores ambientales contribuiría al riesgo final de desarrollar TEA, pero ninguno de estos factores sería determinante por si solo. El resto de casos (aproximadamente 30%) son causados por un **único factor genético determinante**. Este cambio puede afectar a centenares de genes diferentes y puede seguir un patrón de herencia autosómico dominante, autosómico recesivo o recesivo ligada al cromosoma X. En el caso de la **herencia autosómica dominante**, el riesgo

depende de si los padres son portadores o no de la misma alteración genética. En caso de padres portadores, el riesgo de descendencia afecta es del 50%. Si los padres no son portadores, la probabilidad es del 1%, ya que existe el riesgo de mosaicismo germinal, es decir, que la alteración en algunas células reproductivas adicionales (óvulos o espermatozoides) de uno de los dos progenitores. En la **herencia autosómica recesiva**, los dos miembros de la pareja son portadores de una alteración genética y el riesgo de descendencia afecta es del 25%. En el caso de la **herencia recesiva ligada al X**, las mujeres pueden ser portadoras sanas de una alteración. En ese caso, el riesgo para los hijos varones es del 50% y las hijas mujeres tendrían un 50% de riesgo de ser portadoras.

Esta gran heterogeneidad genética dificulta mucho el estudio genético de los TEA. Las pruebas genéticas disponibles en la actualidad en el ámbito clínico solo permiten detectar algunas de las causas determinadas por un único factor genético. Estas pruebas son: cariotipo convencional, cariotipo molecular y estudio molecular del síndrome de X-frágil. Actualmente, permiten encontrar una anomalía genética responsable en el 15% de los casos. En la gran mayoría de casos de TEA, las técnicas actuales no permiten encontrar una alteración genética causante del trastorno.

En los últimos años, se han desarrollado nuevas tecnologías, como la secuenciación de exoma, que permiten estudiar mutaciones puntuales en las regiones codificantes de todos los genes. En la actualidad, estas pruebas sólo se están disponibles en el contexto de estudios de investigación y permiten detectar una anomalía genética responsable en un 20% adicional de casos.

En la Unidad de Genética de la Universitat Pompeu Fabra, se ha analizado mediante cariotipo molecular y secuenciación de exoma una muestra de sangre periférica de PACIENTE, con el fin de ampliar el conocimiento acerca de las bases genéticas del TEA y llegar a encontrar una causa genética que permita dar un mejor asesoramiento genético a las familias. **En el caso de PACIENTE, ambos estudios han sido negativos, y no han permitido encontrar una alteración genética causante de su trastorno.**

En aquellos pacientes en los que no se encuentra una alteración genética responsable del trastorno, el asesoramiento genético se basa en estudios realizados en familias en la misma situación. Estos estudios indican que **el**

riesgo de recurrencia o de tener un siguiente hijo con TEA para los padres que ya han tenido un hijo con TEA es del 10% para autismo clásico y del 20% para rasgos relacionados, como retraso del lenguaje o conductas particulares. El resto de los familiares de una persona afecta de autismo tienen un riesgo superior al de la población general de tener descendencia afecta, aunque no existen datos suficientes que permitan cuantificar este riesgo. La probabilidad de tener un hijo con TEA disminuye a medida que disminuye el grado de parentesco con el familiar afecto. **Cuando no se encuentra una alteración genética responsable de la enfermedad, no es posible aplicar técnicas de diagnóstico prenatal o reproducción asistida que permitan disminuir el riesgo de tener un hijo afecto**.

En cuanto a HERMANO_VARÓN y HERMANA_MUJER, hermanos de PACIENTE, su riesgo de tener descendencia afecta es superior al de la población general y se estima alrededor del 3% y 5% respectivamente. Suponiendo que el TEA que presenta PACIENTE sea de origen multifactorial, el riesgo para ambos sería bajo, ya que la probabilidad que se vuelvan a reunir los mismos factores de riesgo que han actuado en el caso de PACIENTE es baja. En el caso que el TEA que presenta PACIENTE sea causado por un único factor genético determinante, el riesgo depende del patrón de herencia. Si el patrón de herencia es autosómico dominante, el riesgo depende del estatus de portadores de sus padres. En el caso que la alteración genética sea *de novo* y se haya producido por primera vez en PACIENTE, el riesgo para HERMANO_VARÓN y HERMANA_MUJER es igual que al de la población general. En algunos casos de herencia autosómica dominante, las mutaciones pueden tener una penetrancia incompleta, es decir ser heredadas de un progenitor que no presenta el trastorno. La gran mayoría de estas mutaciones son alteraciones del número de copia, que se han descartado mediante cariotipado molecular. En el caso que el TEA que presenta PACIENTE siga un patrón de herencia autosómica recesivo, el riesgo para HERMANO_VARÓN y HERMANA_MUJER depende del estatus de portador de su pareja. Dado que la probabilidad de encontrar una pareja portadora de una alteración en el mismo gen es muy baja, el riesgo en este caso sería próximo a cero. Finalmente, existe la posibilidad de que el TEA que presenta PACIENTE sea debido a una alteración ligada al cromosoma X. En ese caso, el riesgo de descendencia afecta depende del

sexo del progenitor. En el caso de HERMANO_VARÓN, dado que es un varón y no es afecto, no puede ser portador de una alteración ligada al X, por lo tanto su riesgo en este caso sería igual al de la población general. HERMANA_MUJER, en cambio, tendría un 50% de riesgo de ser portadora. En este caso, la probabilidad de tener un hijo afecto sería del 50% si fuera varón. Las hijas de HERMANA_MUJER tendrían un 50% de probabilidades de ser portadoras. Las causas ligadas al X representan sólo un pequeño porcentaje del total y dado que no existe historia familiar que sugiera este patrón de herencia, la probabilidad de que el cuadro de PACIENTE sea ligado al cromosoma X es baja.

En caso de querer recibir más información, se recomienda a los familiares de PACIENTE que acudan a una consulta de asesoramiento genético cuando estén en edad reproductiva. El conocimiento sobre la función de los genes y su papel en enfermedades y trastornos genéticos avanza de forma rápida y progresiva. Por lo tanto, las alteraciones genéticas identificadas y consideradas de significado incierto en la actualidad, deberán ser revisadas más adelante conforme avance el conocimiento científico y se disponga de nueva información que pudiera modificar la interpretación inicial del hallazgo. Si en el transcurso de este proyecto se identificara algún cambio genético considerado causante del trastorno, nos pondríamos en contacto de nuevo con la familia y/o el médico responsable para comunicar los resultados con el fin de proporcionar asesoramiento genético.

Teléfono de contacto: 933160821

Barcelona, 27 de enero de 2016

Fdo.

Dr. Luis Pérez Jurado Dra. Clara Serra Juhé Marta Codina Solà

DISCUSSION

Exploring ASD genetic architecture through high-throughput sequencing

Our understanding of ASD genetic architecture has changed dramatically over the last years. Twin studies have long showed that genetics plays a major role in the etiology of ASD, but the identification of specific genes has proven more difficult (36–43). In the recent years, new developments in genomic technology have enabled the identification of a large number of new loci. First, the advent of CMA allowed the identification of small rearrangements in a genome-wide manner, uncovering the role of rare CNVs and recurrent microdeletion and microduplication events (53–56). More recently, the ability to sequence the entire genome or sets of protein-coding genes in a rapid and cost-effective manner has enabled the study of both common and rare genetic variants (66–76). The results of these recent approaches have uncovered the great genetic heterogeneity behind the wide range of expressivity of the disorder. These advances highlight the notion that ASD is not a single entity, but an ensemble of conditions that share a set of common features, just as its name implies.

In this work, we have applied various NGS based strategies to study the contribution of rare and complex variants to the genetic etiology of ASD. First, we have integrated exome and transcriptome sequencing to explore the role of rare variation and its expression consequences. Selected loci have been sequenced in an extended cohort of cases and controls to confirm their pathogenic role. In addition, we have explored the role of second-hit variants in ASD susceptibility acting in the presence of a major genetic lesion, the WBS deletion, as well as the contribution of complex genomic variants to the etiology of ASD. Finally, to study if the recent advances in genetics have reached society, as well as to contribute to the knowledge translation of research findings, we have assessed parents' knowledge and perceptions regarding ASD and provided genetic counseling to all interested families.

The combined use of these different approaches has contributed to a better understanding of the interplay of genetic factors and their contribution to the liability and variable expressivity of ASD. Throughout this study, we have identified examples of various models of inheritance, ranging from monogenic cases, with individuals carrying highly penetrant variants, to individuals

harboring a combination of inherited rare variants of milder penetrance, probably conforming an oligogenic model. Despite being found in a wide range of loci, affected genes converge in several functional pathways, such as cell adhesion molecules involved in synapse development and the PI3K/Akt intracellular signaling pathway. In addition, this thesis has contributed to the diagnosis of several cases formerly considered idiopathic, as well as to gain a better insight of inheritance patterns in families with no identifiable cause, resulting in genetic counseling for participating families. Our results show an underutilization of genetic services among parents, which have an adverse impact on access to genetic services and recurrence risk estimation. In addition, we show that genetic counseling has positive effects on knowledge and recurrence risk estimates.

Highly penetrant variants

Over the last years, several studies have shown that *de novo* mutations play an important role to the causality of ASD and are estimated to contribute in about 20% of cases, with *de novo* LoF mutations present in approximately 10% of cases and missense variants contributing to an additional 10% of patients (102). Our first study integrating exome and transcriptome sequencing in 36 males with idiopathic ASD revealed *de novo* mutations in a similar proportion of individuals, approximately 14%. Besides *de novo* mutations, highly penetrant variants following and X-linked pattern of inheritance were found in two additional cases, representing around 5% of individuals. Even though our study was not trio-based, our approach resulted in similar estimates of contributing *de novo* mutations, validating our prioritization strategy based on the integration of omics data. The study of the transcriptional consequences of genetic variants by transcriptome sequencing also allowed the identification of an additional case that would have been missed using standard filtering pipelines. Moreover, we identified several genes with rare variants showing nonsense-mediated decay, allele specific expression or changes in expression levels.

Highly penetrant variants were identified in nine different loci: *PTEN*, *MAOA*, *SCN2A*, *CDKL5*, *SHANK3*, *CNTNAP3*, *CUL3*, *MED13L* and *KCNV1* (Figure 1). Some of these loci have already been reported before and their association to ASD has been extensively proven, whereas for others none or very few cases had been previously described. The identification of variants

in known loci highlights the relevance of genetic diagnosis in ASD, since some of the results have implications for clinical management or increased recurrence risk. For instance, a patient with a *de novo* pathogenic mutation in *PTEN* was identified during the course of this study, a finding with relevant clinical implications. Mutations in *PTEN* have been described in approximately 20% of patients with ASD and macrocephaly (334,335). In addition, they give rise to a variety of syndromes with increased risk of hamartomatous tumors, which are grouped together under the term PHTS (336). Functional analysis of *PTEN* mutations suggests that variants resulting in neurodevelopmental disorders or increased tumor susceptibility may have different protein effects. Whereas mutations that partially preserve protein activity may give rise to neurodevelopmental phenotypes, variants completely disrupting protein activity may result in increased tumor risk (337,338). However, several authors have suggested that all patients with *PTEN* germline mutations should follow the same surveillance protocol for malignant tumors (339–341). It is important to note that this mutation was identified thanks to the joint analysis of exome and transcriptome sequencing, showing the potential of transcriptome sequencing when applied to clinical diagnostic. Since the variant identified lied in an intronic region, it would have been missed by standard filtering, even though it activated a cryptic splice site and resulted in an aberrant transcript.

Besides *PTEN*, we have identified variants in two X-linked genes (*MAOA* and *CDKL5*), with important familial implications, since female carriers have a 50% chance of transmitting the mutated allele. A splicing variant in *MAOA* was found in two affected brothers and his mother, all presenting altered levels of urinary monoamines (342). At the time of publication, this was the third family described since 1993 with mutations in this gene (343,344). Since then, two more cases have been described, leading to potential dietary and therapeutic recommendations that could ameliorate the patients' symptoms (345). Additionally, we identified two patients with variants affecting another X-linked gene, *CDKL5.* The first variant was found in a male with an Angelman-like phenotype. Although it was inherited from his mother and was also present in his unaffected sister, the selective inactivation of the mutated allele in the sister could explain their unaffected status. The second variant was a *de novo* missense variant found in a female patient, with a phenotypic presentation reminiscent of previous cases, which are beginning to delineate

a separate entity (346). This suggests that both *de novo* damaging variants with a dominant effect in females, as well as hypomorphic alleles acting in a recessive way in males play a role in the etiology of ASD.

Likewise *CDKL5*, recurrent mutations have also been identified in *SCN2A,* with two patients harboring nonsense mutations in this locus: one *de novo* and one with unknown inheritance due to the lack of available parental samples. *SCN2A* is one the few genes with recurrent LoF *de novo* mutations and, strikingly one of the loci with the highest number of variants described (68–70,74,76,347,348). In addition to *SCN2A*, we also identified *de novo* variants in *SHANK3* and *CNTNAP2*, two genes belonging to the neurexin superfamily of cell adhesion molecules, crucial for synaptic function and connectivity. Interestingly, most mutations affecting *CNTNAP2* reported before are inherited from unaffected parents and very few *de novo* variants have been described in patients with ASD (60,349). This fact highlights the complex genetic architecture of ASD, where mutations in the same gene can act following an autosomal recessive or dominant model. In the latter case, both inherited mutation with incomplete penetrance as well as *de novo* mutations of complete penetrance perhaps acting in a sensitized background would play a role.

In addition to established candidate genes, *de novo* variants were also identified in genes were fewer *de novo* variants had been described at the time, such as *CUL3* and *MED13L.* The additional cases identified in this thesis will contribute to the phenotypic characterization of the patients, which will hopefully result in the definition of specific ASD subtypes and the development of tailored management and therapeutic strategies. Finally, our work points towards a new candidate locus for ASD, since we identified a patient with a *de novo* LoF mutation in a *KCNV1*. *KCNV1* encodes for a potassium channel subunit highly expressed in brain with a high probability of being intolerant to LoF variants, making it a great candidate gene for ASD. Since the identification of a single LoF *de novo* variant does not imply the functional involvement of a gene, further functional studies and/or in larger cohorts are needed to confirm or discard its role. To this end, large sequencing projects integrating phenotypic and genetic data, such as SPARK

(Simons Foundation Powering Autism Research for Knowledge, https://sparkforautism.org/portal/homepage/) will be key to achieve this aim.

Therefore, our work reinforces the role of highly penetrant genes in the pathogenesis of ASD, showing that *de novo* SNVs could explain approximately 19% of cases so far considered idiopathic. However, the results also suggest the contribution of inherited rare variants following an oligogenic model in some additional cases that could explain an additional part of the missing heritability in ASD.

The contribution of inherited rare variants to oligogenic models

Besides *de novo* variants, the contribution of rare inherited variants is also important as shown by recent studies relying on larger samples (74). Since this type of variation is inherited from unaffected or mildly affected progenitors, its individual effect would not be sufficient to cause the disorder. However, they could act as phenotypic modifying factors that create a sensitized background that only manifests in the presence of a major genetic lesion. This would explain the wide range of expressivity seen in disorders of full penetrance, as well as the incomplete penetrance and wide expressivity of some CNVs. On the other hand, the accumulation of rare variants could also be pathogenic by itself. In that case, each parent would carry a few rare events of medium penetrance, and ASD would only result from the combination of factors inherited from both lines, consistent with an oligogenic model. Our work provides specific examples of the role of inherited variants in both models (Figure 2).

First, inherited second-hit variants acting as phenotypic modifiers were detected in two different disorders of full penetrance: disruptions of *SHANK3* and WBS. By targeted sequencing of candidate genes, we detected an individual with a *de novo* nonsense mutation affecting *SHANK3*, who also carried a maternally inherited nonsense variant in *CNTN6*. As previously mentioned, *SHANK3* encodes a protein located at the postsynaptic density that connects various membrane proteins to the actin cytoskeleton and mediates signaling in various functional pathways, such as the Pi3K/Akt (Figure 1) (350). *SHANK3* is also deleted in Phelan-McDermid syndrome caused by a loss of 22q13 (351). Phenotypic variability has been observed both in individuals with deletions of 22q13 and patients carrying *de novo*

mutations, suggesting that other genetic factors, besides the range in deletion size, are responsible for the variable expressivity (352,353). In this regard, a previous case of an individual with a *de novo* mutation in *SHANK3* and an additional double-hit CNV has been described (354,355). Interestingly, *CNTN6* has been proposed before as a second-hit factor in a girl with a 16p11.2 deletion (356). CNVs in *CNTN6*, both inherited and *de novo*, have been described in patients with neurodevelopmental disorders, although the frequency in cases does not seem to be significantly increased (357,358). Taken together, these evidences suggest that variants altering *CNTN6* could create a sensitized genetic background and act as modifying phenotypic factors in the presence of a major genetic lesion.

Besides susceptibility variants acting in a background of genetic lesions disrupting ASD candidate genes, second-hit variants could also predispose to ASD in a major genetic lesion resulting in a mirror phenotype: the WBS. Although most individuals with WBS show a particular neurocognitive profile with an overly friendly and empathetic personality, around 12% of patients present ASD features (270). Therefore, the prevalence of ASD among WBS individuals is increased by 10x compared to that in the general population, which is striking considering the typical neurocognitive profile associated with WBS. Our study of second-hit modifying variants in eight individuals with WBS and co-occurring ASD suggests that rare variants in highly constrained brain expressed genes could play a role in the phenotypic variability observed in these patients. Our unbiased approach using exome sequencing identified LoF and deleterious variants in candidate genes previously associated to ASD. Interestingly, females presented a higher burden of rare mutations in candidate genes. This phenomenon has been widely described in ASD, with *de novo* LoF variants being more frequent in affected females than affected males and the male to female ratio being more biased in high functioning cases (25,66–70). This is consistent with a female protective effect, in which females would need a higher mutational burden in order to develop the disorder. Our findings suggest that the female protective effect may still play a role in disorders of full penetrance in which ASD is not one of the common features.

Moreover, the study of second-hit SNVs in disorders of full penetrance, but variable expressivity, can help uncover new genes with milder effects that

manifest only in a sensitized background. For example, our analysis of second-hit variants in individuals with WBS and co-ocurring ASD also revealed LoF variants in highly constrained genes. Interestingly, all genes with LoF in our cohort selected on the bases of being intolerant were brain-expressed genes. Some of the genes (*PRR12*, *CXXC1* and *SEC24C)* had also been previously associated to various neuropsychiatric disorders sharing a common neurobiological and genetic basis with ASD (ID, bipolar disorder and schizophrenia, respectively) (359–361). Such variants could also contribute to ASD susceptibility in other cases without a major hit, following an oligogenic model. Individually, each variant would not be sufficient to cause the disorder, but would contribute to increase risk. Since each factor would exert a smaller effect, their detection would be more difficult.

Besides variation acting as a second-hit in the presence of a major genetic alteration, we have also explored the role of rare inherited variants in patients with idiopathic ASD. First, our study integrating exome and transcriptome sequencing revealed that inherited rare inherited variants resulted in common functional consequences. Enrichment analysis of individual rare variation revealed functional pathways exclusively overrepresented in ASD patients, such as the Pi3K-Akt signaling and axon guidance pathways (Figure 1). The incorporation of peripheral blood transcriptome data provided additional information to detect genetic variants associated with functional consequences. For instance, we identified a male patient carrying a rare variant in *MECP2* who presented overexpression of the same gene, as well as a patient carrying three rare mutations in candidate genes *ANK3*, *CREBBP* and *SEMA6B* with concurring overexpression of the same loci. Overexpression of these loci could be due to a positive feed-back mechanism, which would upregulate the expression genes with impaired function due to the presence of the identified mutations.

The contribution of rare variants following an oligogenic model is also supported by the identified specific examples of variants cosegregating with neuropsychiatric phenotypes in some families. For example, a novel damaging missense variant in *NRXN2* was identified in two affected brothers and was found to be inherited from his father, presenting OCD, a neuropsychiatric disorder with a shared neurobiological basis with ASD. Interestingly, a female sibling of the family who did not present the damaging

variant had been also diagnosed with Specific Language Impairment (SLI) (Figure 2). This suggests the presence of additional genetic factors running in the family, which may be shared among affected individuals, but also present in unaffected relatives or relatives with milder related phenotypes. In addition to the previously mentioned variant in *NRXN2,* we identified several other examples of inherited pathogenic variants in *CNTNAP2*, specifically a nonsense mutation and a missense damaging altering protein traffic, in parents with no obvious phenotypes. To further support the role of the neurexin superfamily of genes in ASD, a case-control study revealed a significant increase in the burden of rare mutations in synapse related genes (Figure 1). However, a detailed breakdown revealed that this association was driven only by LoF intolerant genes, in agreement with previous studies showing that ASD candidate genes are enriched in functionally constrained genes (362). Nonetheless, that finding does not undermine the potential role of LoF tolerant genes in ASD, but rather suggests that their effect may be smaller and that larger cohorts are required to assess their possible contribution. Finally, this result illustrates the advantages of pathway-based approaches compared to gene-based case-control association studies. Enrichment analyses based on functional categories avoid the use of multiple-test corrections while relying on functional evidence (363). This makes them an approach especially well suited for complex disorders of high genetic heterogeneity, such as ASD.

Despite this high degree of genetic heterogeneity, previous studies have shown that mutated loci in ASD converge in a set of functionally related pathways (76). Similarly, altered genes in our cohort cluster in functionally related networks, mainly associated with synapsis and the Pi3K/Akt signaling pathway (Figure 1). Highly penetrant mutations have been observed in various genes from the neurexin superfamily, as well as in voltage-gated ion channels and neurotransmitter regulators. In addition, we also found that rare inherited variants are enriched in the Pi3K/Akt signaling pathways in ASD patients, but not in controls. This is consistent with an epistatic oligogenic model where inherited variation disrupts a common signaling pathway. The Pi3K pathway regulates the mammalian target of rapamycin (mTOR) signaling pathway, a major regulator of cell growth. The mTOR pathway is deregulated in patients with mutations in *NF1*, *TSC1* or *PTEN*, which was also mutated in our cohort (103).

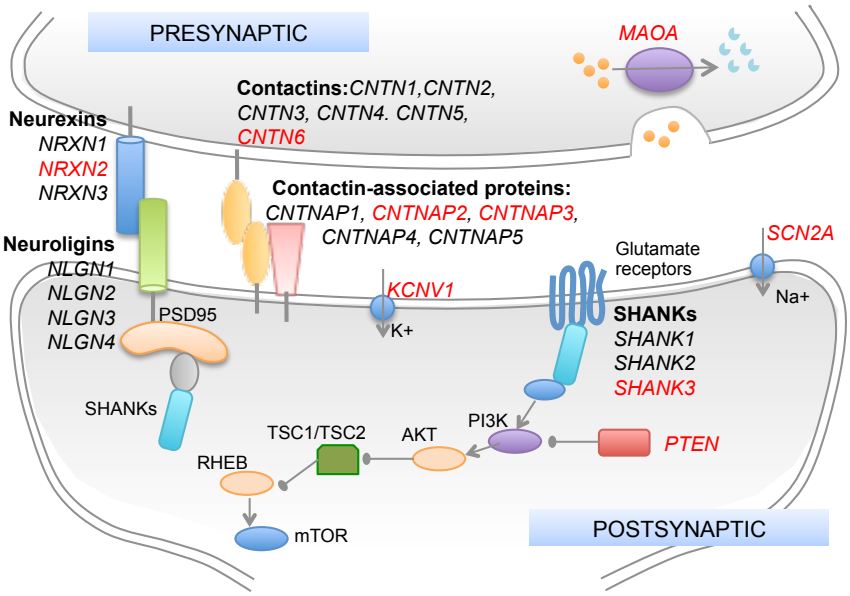


Figure 1. Mutated loci in our cohort converge in two functionally related pathways, associated with synapse function and the Pi3K/Akt signaling pathway.

Mutations in recently duplicated genes

In addition to rare SNVs, part of the missing heritability of ASD could also be lying in complex regions of the genome that have been poorly explored until now, such as segmental duplications (SD). The fact that many SD have appeared specifically in the human genome supports their contribution to human behavior (178). In addition, their complex structure predisposes them to rearrangements. Two of these loci are the 7q11.23 region and the pericentromeric region of chromosome 9, both containing candidate brain expressed genes (*GTF2I*, *GTF2IRD2* and *CNTNAP3*). Despite their interest, these regions remain largely unstudied due to their inherent difficulty. To study the variation and the possible contribution of these two regions to ASD, we also performed an exploratory study based on targeted sequencing and developed a pipeline to identify CNVs and SNVs. Our strategy, based on the quantitative nature of read-depth and the specific copy information provided by PSVs, successfully identified rearrangements in both regions.

Although rearrangements were identified and validated in both regions, the previous knowledge about the detailed architecture of the 7q11.23 region

allowed a much finer characterization of the identified CNVs. The results in this region revealed an increased frequency of duplications containing specifically the B block in cases compared to controls. Although a further characterization of the origin and breakpoint of the rearrangements is needed, the results point towards a role of *GTF2IRD2*. As mentioned in previous chapters, the B block contains three genes: *GTF2I*, *GTF2IRD2* and *NCF1.* Since part of *GTF2I* lies in the single copy region not included in the identified rearrangements, none of the duplications contain a functional copy of this gene. Therefore, due to its role as a possible modulator of the phenotype in WBS, *GTF2IRD2* seems the most plausible candidate. However, the effect of the duplications in *GTF2IRD2* gene dosage is unknown. Although duplications are expected to result in increased gene dosage, the finer characterization of the rearrangements by experimental methods suggests that they may result in the disruption of a functional copy of *GTF2IRD2*. Hence, a detailed characterization of the duplication breakpoint is warranted. Other possible explanations include positional effects altering the expression of nearby genes. The role of *GTF2IRD2* is also supported by an additional pathogenic mutation affecting the start methionine codon of this gene in two affected brothers and their unaffected father. *In silico* prediction of alternatives reading frames suggests the usage of an alternate methionine codon located nearby. Additional studies of protein function will help to elucidate the consequences of this variant. Taken together these results suggest that rearrangements in the 7q11.23 region may increase susceptibility to ASD. Since the same type of rearrangement has also been observed in controls, albeit at a lower frequency, they may act as susceptibility variants in the context of a polygenic model.

Regarding the pericentromeric region of chromosome 9, our results show a much higher level of polymorphic variation and gene conversion events, that difficult even more its study. Despite this, our tailored approach based on measuring ZRPKM on windows overlapping with partial copies has enabled the accurate measurement of the total copy number of partial and functional copies of *CNTNAP3*. Our approach is validated by comparison with previous results showing similar and accurate distributions of copy number in the region. In addition, the detailed study of the sequence and its expression in PBMCs and fetal brain has enabled the definition of specific PSVs and the design of experimental methodologies for validation and extended study.

Overall, the results do not reveal a difference in the frequency of rearrangements or SNVs between cases and controls. However, some of the findings suggest functional alterations of *CNTNAP3B* in ASD patients. First, one of the deletions identified in patients was found to be *de novo*. In addition, we identified an aberrant pattern of expression in one of the patients, who expressed both *CNTNAP3A* and *CNTNAP3B* in blood, also present in his mother. Thus, further studies of the region in a larger cohort will help elucidate the possible role of this candidate region in the disorder.

In summary, our results illustrate the complex genetic nature of ASD. We have identified cases following various models of inheritance, with different contributions of de novo and inherited rare variants (Figure 2). In addition, the data confirm the genetic heterogeneity of ASD, which is reflected in the highly variable expressivity.

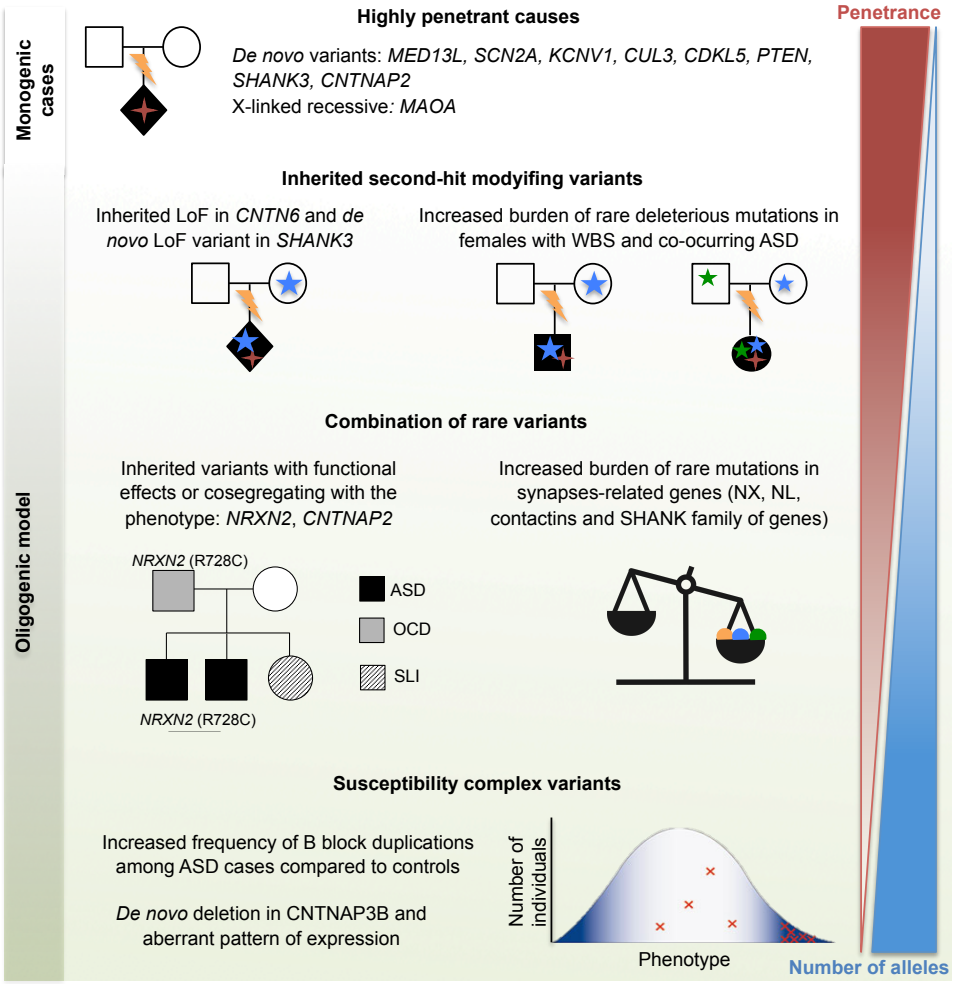


Figure 2. Summary of the main results of this thesis, ordered from top to bottom, according to the penetrance and polygenic nature of genetic factors involved.

Genetic counseling in idiopathic ASD in the genomics era

As we have seen in the previous section, the genetic architecture of ASD is extremely complex, includes various models of inheritance and carries important familial implications. Moreover, despite the tremendous advances of the last decade, most cases remain idiopathic. Therefore, families must deal with a great amount of information and understand technical concepts in order to take informed decisions. Besides the complexity of the information provided, taking care of a child with ASD has a great impact on family dynamics. For all these reasons, parents would benefit greatly from genetic counseling, which can provide accurate and understandable information, as well as psychological support to families.

To explore the utility and extend of genetic counseling in ASD, we have evaluated the utilization of genetic services, perceived causes, knowledge and opinion about genetics in a group of parents with children with idiopathic ASD. The results show a widespread underutilization of genetic services. In fact, none of the participants recruited through family associations had seen a clinical geneticist, in contrast with current guidelines for diagnostic evaluation. It is also striking that, despite the clear underutilization of genetic services, 27% of participants recruited through families associations and 60% of the total sample of parents had undergone some type of genetic testing, mainly karyotype and X-fragile testing. Therefore, most parents had not access to CMA, which is considered the clinical standard for first tier genetic testing in ASD. The underutilization of genetic services may also be the reason behind the lack of knowledge about practical aspects of genetic testing in the studied group of parents. Parents were asked several questions about the meaning of genetic results, such as if negative results meant that recurrence risk in a hypothetic subsequent pregnancy was similar to that of the general population. For most questions, most participants did not know the answer or selected an incorrect option.

To examine the reasons behind the underutilization of genetic services, parents were asked about their interest in further genetic testing. An overwhelming majority of parents answered that they would be interested in genetic testing, mainly to improve the medical care of their affected children. This fact may be related to parental age, as well as to the time elapsed between the diagnosis of their affected child and the present study. Although

information for a next pregnancy was not one of the main reasons for interest in genetic testing, half of the participants admitted that recurrence risk had an effect on family planning. In fact, recurrence risk which was highly overestimated by most participants. Numerical estimates of risk correlated with categorical estimates, with most participants perceiving their risk as high" or "very high".

Finally, our study showed the value of genetic counseling in helping parents deal with the great amount of complex information surrounding ASD. A group of participants who underwent CMA and exome sequencing were offered genetic counseling sessions. The results show that knowledge and numerical recurrence risk estimates improved both after sessions and fifteen days later. Not only numerical risk estimates were more accurate after sessions, but also qualitative estimates of perceived recurrence risk reduced for most participants. However, perceived personal control measurements did not change, in agreement with previous studies showing higher increments for participants with positive results (364). Last, and beyond the numbers, sessions also provided parents and close relatives with a space to share their experiences, worries and feelings. One of the most recalled experiences was the initial confusion after the initial diagnosis, which for most of the parents had taken place years ago. Most participants stated that they would have welcomed professional guidance, as well as an integral care plan and hoped that the situation would have improved in the last years. In some families, siblings also attended counseling sessions, revealing the impact that caring for a child on the spectrum has on family dynamics. Some expressed their worries about the responsibility that entailed taking care of their affected siblings, as well as about their own recurrence risk.

In summary, the results of this work offer a glimpse into parental perspectives of ASD. Similar to previous studies, very few parents had seen a genetic professional, despite genetics being perceived as the most prevalent cause. Risk recurrence was highly overestimated and had a great influence on family planning. Finally, our results also show that genetic counseling can help parents improve their knowledge about familial implications, as well as to better estimate their recurrence risk. Therefore, this work underscores the need for genetic testing and genetic counseling in individuals with ASD as part of an integrated care program, tailored according to individual needs.

Concluding remarks

ASD are highly heterogeneous disorders, both from a genetic and clinical point of view. This great genetic heterogeneity hinders its methodological approach, although a few strategies can help overcome this limitation. First, it is essential to have a detailed clinical characterization of patients, which enables the identification of ASD subtypes. Obtaining a detailed family history, as well as relative samples, can also help identify mildly affected relatives and inheritance patterns. Finally, the study of the functional consequences of variation can help identify pathogenic variants and prove its casualty.

By using these strategies, we have identified several cases following different inheritance patterns, showing the complex genetic architecture of ASD. Our data points to the contribution of highly penetrant variants, following a *de novo* or X-linked inheritance in approximately 20% of cases. Several evidences also indicate the contribution of variants of milder penetrance in oligogenic models, as well as phenotypic modifying factors. This complexity poses a great challenge for genetic diagnosis and makes genetic counseling essential for affected families. The work presented in this thesis will contribute to a better understanding of ASD genetic architecture. The discovery of new loci will improve genetic diagnosis and genetic counseling, by providing families with a specific recurrence risk. Ultimately, we hope that the identification of convergent pathways results in personalized therapeutic approaches that will improve the well-being of affected individuals.

CONCLUSIONS

Conclusions

1. The joint study of genomic and transcriptomic sequencing is a good strategy to identify genetic variants related to ASD and has allowed the identification of additional monogenic causes, as well as of several pathogenic variants that would have been missed by standard filtering.

2. Highly penetrant genetic variants were found in 19% of the patients in our study by exome sequencing, being 14% *de novo* and 5% X-linked. Variants were found in seven different genes, *SCN2A*, *CDKL5*, *PTEN*, *MAOA*, *CUL3*, *KCNV1* and *MED13L*, and no recurrent hits were identified.

3. The vast majority of the genetic variants identified by exome sequencing were inherited from unaffected progenitors, implying that they could not be the sole cause of the disorder. Nonetheless, some could contribute to its susceptibility as part of an oligogenic or multifactorial model. This is supported by enrichment analyses showing their accumulation in two pathways related to the disorder: axon guidance and Pi3K/Akt signaling.

4. Sequencing of an additional cohort of cases and controls revealed additional *de novo* variants in *SCN2A* and *CDKL5*, suggesting that some loci may be more frequently mutated than others and highlighting the extreme heterogeneity of the disorder.

5. Synapse related genes play a major role in the etiology of ASD. *De novo* pathogenic mutations were identified in *SHANK3* and *CNTNAP2*, and variants cosegregating with the phenotype in *CNTNAP2* and *NRXN2*. In addition, affected individuals carried a significantly higher burden of rare variants in LoF intolerant genes, compatible with an oligogenic model of the disorder.

6. Our sequencing approach and adapted pipeline targeting duplicated genes can successfully detect rearrangements and point mutations in complex regions of the genome.

7. Rearrangements containing only B block in 7q11.23 were significantly more frequent in cases than controls, pointing to their contribution as susceptibility factors to ASD and a possible role of *GTF2IRD2*.

8. The pericentromeric region of chromosome 9 harboring *CNTNAP3* is highly polymorphic, with most variation being due to partial non-functional copies. The frequency of rearrangements affecting functional copies did not differ significantly between cases and controls, although a *de novo* deletion and an aberrant pattern of expression were detected in two cases.

9. The search for second-hit genetic factors in a group of individuals with WBS and co-occurring ASD did not reveal a contribution of *cis* or *trans* acting factors in the remaining hemizygous allele, neither of rare CNVs or common genetic variants. All deletions were of paternal origin, suggesting a role of epigenetic control mechanisms.

10. Individuals with WBS and co-occurring ASD presented several LoF variants in highly constrained genes, including previously known candidate genes. Interestingly, females carried a higher burden of rare deleterious variants in candidate genes compared to their male counterparts, in agreement with the well established female protective effect in ASD.

11. Despite the strong genetic component of ASD, we found an extended underutilization of genetic services among families. Since the most frequently perceived cause of ASD was genetics and parents' interest in genetic testing was high, this lack of access could not explained by families' disinterest. The underutilization of genetic services has relevant implications for family planning, since most of patients do not receive the recommended first-tier diagnostic tests and parents highly overestimate their recurrence risk.

12. Our results show that genetic counseling can help parents improve their knowledge about the genetic implications of ASD, as well as to better estimate their recurrence risk and provide psychological support to families. Genetic counseling should be included in integrated care programs, tailored according to each individual needs.

REFERENCES

References

1. Kanner L. Autistic disturbances of affective contact. Nervous Child. 1943. p. 217–50.

2. Asperger H. Die „Autistischen Psychopathen" im Kindesalter. Arch Psychiatr Nervenkr. 1944 Jun;117(1):76–136.

3. Zwaigenbaum L, Bryson S, Garon N. Early identification of autism spectrum disorders. Behav Brain Res. 2013;251:133–46.

4. Barger BD, Campbell JM, McDonough JD. Prevalence and onset of regression within autism spectrum disorders: A meta-analytic review. J Autism Dev Disord. 2013;43(4):817–28.

5. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. American Psychiatric Association. 2013. 991 p.

6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Text. 2000. 943 p.

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

8. Matson JL, Shoemaker M. Intellectual disability and its relationship to autism spectrum disorders. Res Dev Disabil. 2009 Jan;30(6):1107–14.

9. Spence SJ, Schneider MT. The role of epilepsy and epileptiform EEGs in autism spectrum disorders. Pediatr Res. 2009 Jun;65(6):599–606.

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

11. Valicenti-McDermott M, McVicar K, Rapin I, Wershil BK, Cohen H, Shinnar S. Frequency of gastrointestinal symptoms in children with autistic spectrum disorders and association with family history of autoimmune disease. J Dev Behav Pediatr. 2006;27(2):128–S136.

12. Antshel KM, Zhang-James Y, Faraone S V. The comorbidity of ADHD and autism spectrum disorder. Expert Rev Neurother. 2013 Oct;13(10):1117–28.

13. Matson JL, Williams LW. Depression and mood disorders among persons with autism spectrum disorders. Res Dev Disabil. 2014 Sep;35(9):2003–7.

14. Miles JH, Hadden LL, Takahashi TN, Hillman RE. Head circumference is an independent clinical finding associated with autism. Am J Med Genet. 2000 Dec 11;95(4):339–50.

15. Fombonne E, Rogé B, Claverie J, Courty S, Frémolle J. Microcephaly and macrocephaly in autism. J Autism Dev Disord. 1999 Apr;29(2):113–9.

16. Ozgen HM, Hop JW, Hox JJ, Beemer FA, van Engeland H. Minor physical anomalies in autism: a meta-analysis. Mol Psychiatry. 2010 Mar;15(3):300–7.

17. Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. MMWR Surveill Summ. 2014 Mar 28;63(2):1–21.

18. Shattuck PT. The contribution of diagnostic substitution to the growing administrative prevalence of autism in US special education. Pediatrics. 2006 Apr;117(4):1028–37.

19. Polyak A, Kubina RM, Girirajan S. Comorbidity of intellectual disability confounds ascertainment of autism: implications for genetic diagnosis. Am J Med Genet Part B Neuropsychiatr Genet. 2015 Oct 22;168(7):600–8.

20. Taylor B, Jick H, Maclaughlin D. Prevalence and incidence rates of autism in the UK: time trend from 2004-2010 in children aged 8 years. BMJ Open. 2013;3(10):e003219.

21. Kaye JA, del Mar Melero-Montes M, Jick H. Mumps, measles, and rubella vaccine and the incidence of autism recorded by general practitioners: a time trend analysis. BMJ. 2001 Feb 24;322(7284):460–3.

22. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2000 Principal Investigators, Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, six sites, United States, 2000. MMWR Surveill Summ. 2007 Feb 9;56(1):1–11.

23. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2002 Principal Investigators, Centers for Disease Control and Prevention. Prevalence of autism spectrum disorders--autism and developmental disabilities monitoring network, 14 sites, United States, 2002. MMWR Surveill Summ. 2007 Feb 9;56(1):12–28.

24. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2006 Principal Investigators, Centers for Disease Control and Prevention (CDC). Prevalence of autism spectrum disorders - Autism and Developmental Disabilities Monitoring Network, United States, 2006. MMWR Surveill Summ. 2009 Dec 18;58(10):1–20.

25. Autism and Developmental Disabilities Monitoring Network Surveillance Year 2008 Principal Investigators, Centers for Disease Control and Prevention, Baio J et al. Prevalence of autism spectrum disorders--Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. MMWR Surveill Summ. 2012 Mar 30;61(3):1–19.

26. WHO. International Statistical Classification of Diseases and Related Health Problems (International Classification of Diseases)(ICD) 10th Revision - Version:2010. Occupational Health. 2010. 1-201 p.

27. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition. Text. 1987.

28. First MB. Diagnostic and Statistical Manual of Mental Disorders DSM IV. Professional Psychology Research and Practice. 1994. 465-473 p.

29. American PSychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV-TR. 4th edition. 2000.

30. Zander E, Bölte S. The New DSM-5 Impairment Criterion: A Challenge to Early Autism Spectrum Disorder Diagnosis? J Autism Dev Disord. 2015 Nov;45(11):3634–43.

31. Ohashi K, Mizuno Y, Miyachi T, Asai T, Imaeda M, Saitoh S. Concordance of classifications using DSM-5 and DSM-IV-TR criteria for autism spectrum disorder. Pediatr Int. 2015 May 25;

32. Smith IC, Reichow B, Volkmar FR. The Effects of DSM-5 Criteria on Number of Individuals Diagnosed with Autism Spectrum Disorder: A Systematic Review. J Autism Dev Disord. 2015 Aug;45(8):2541–52.

33. Kulage KM, Smaldone AM, Cohn EG. How will DSM-5 affect autism diagnosis? A systematic literature review and meta-analysis. J Autism Dev Disord. 2014 Aug;44(8):1918–32.

34. Maenner MJ, Rice CE, Arneson CL, Cunniff C, Schieve LA, Carpenter LA, et al. Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates. JAMA psychiatry. 2014 Mar;71(3):292–300.

35. Coury DL. DSM-5 and autism spectrum disorders: implications for families and clinicians. J Dev Behav Pediatr. 2013 Sep;34(7):494–6.

36. Folstein S, Rutter M. Infantile autism: a genetic study of 21 twin pairs. J Child Psychol Psychiatry. 1977 Sep;18(4):297–321.

37. Ritvo ER, Freeman BJ, Mason-Brothers A, Mo A, Ritvo AM. Concordance for the syndrome of autism in 40 pairs of afflicted twins. Am J Psychiatry. 1985 Jan;142(1):74–7.

38. Steffenburg S, Gillberg C, Hellgren L, Andersson L, Gillberg IC, Jakobsson G, et al. A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. J Child Psychol Psychiatry. 1989 May;30(3):405–16.

39. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, Yuzda E, et al. Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med. 1995 Jan;25(1):63–77.

40. Taniai H, Nishiyama T, Miyachi T, Imaeda M, Sumi S. Genetic influences on the broad spectrum of autism: study of proband-ascertained twins. Am J Med Genet B Neuropsychiatr Genet. 2008 Sep 5;147B(6):844–9.

41. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. Arch Pediatr Adolesc Med. 2009 Oct;163(10):907–14.

42. Lichtenstein P, Carlström E, Råstam M, Gillberg C, Anckarsäter H. The genetics of autism spectrum disorders and related neuropsychiatric disorders in childhood. Am J Psychiatry. 2010 Nov;167(11):1357–63.

43. Ronald A, Hoekstra R a. Autism spectrum disorders and autistic traits: A decade of new twin studies. Am J Med Genet Part B Neuropsychiatr Genet. 2011;156(3):255–74.

44. Ozonoff S, Young GS, Carter A, Messinger D, Yirmiya N, Zwaigenbaum L, et al. Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics. 2011 Sep;128(3):e488–95.

45. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. The familial risk of autism. JAMA. 2014 May 7;311(17):1770–7.

46. Devlin B, Scherer SW. Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev. 2012 Jun;22(3):229–37.

47. Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. Brain Res. 2011 Mar 22;1380:42–77.

48. Miles JH, McCathren RB, Stichter J, Shinawi M. Autism Spectrum Disorders. University of Washington, Seattle; 2010.

49. Kent L, Evans J, Paul M, Sharp M. Comorbidity of autistic spectrum disorders in children with Down syndrome. Dev Med Child Neurol. 1999 Mar;41(3):153–8.

50. van Rijn S, Stockmann L, Borghgraef M, Bruining H, van Ravenswaaij-Arts C, Govaerts L, et al. The social behavioral phenotype in boys and girls with an extra X chromosome (Klinefelter syndrome and Trisomy X): a comparison with autism spectrum disorder. J Autism Dev Disord. 2014 Feb;44(2):310–20.

51. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-scale copy number polymorphism in the human genome. Science. 2004 Jul 23;305(5683):525–8.

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

53. Cuscó I, Medrano A, Gener B, Vilardell M, Gallastegui F, Villa O, et al. Autism-specific copy number variants further implicate the phosphatidylinositol signaling pathway and the glutamatergic synapse in the etiology of the disorder. Hum Mol Genet. 2009 May 15;18(10):1795–804.

54. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. Nature Publishing Group; 2009 May 28;459(7246):569–73.

55. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. J Hum Genet. 2008;(February):477–88.

56. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science. 2007 Apr 20;316(5823):445–9.

57. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron. 2011 Jun 9;70(5):898–907.

58. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron. 2011 Jun 9;70(5):863–85.

59. Rosenfeld J a, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. Genet Med. 2013;15(6):478–81.

60. Gregor A, Albrecht B, Bader I, Bijlsma EK, Ekici AB, Engels H, et al. Expanding the clinical spectrum associated with defects in CNTNAP2 and NRXN1. BMC Med Genet. 2011 Jan;12:106.

61. D'Angelo D, Lebon S, Chen Q, Martin-Brevet S, Snyder LG, Hippolyte L, et al. Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. JAMA psychiatry. 2015 Dec 2;1–11.

62. Shinawi M, Liu P, Kang S-HL, Shen J, Belmont JW, Scott D a, et al. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. J Med Genet. 2010;47:332–41.

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

64. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature. 2014;505(7483):361–6.

65. Hanson E, Bernier R, Porche K, Jackson FI, Goin-Kochel RP, Snyder LG, et al. The cognitive and behavioral phenotype of the 16p11.2 deletion in a clinically ascertained population. Biol Psychiatry. 2015 May 1;77(9):785–93.

66. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature. Nature Publishing Group; 2012 Apr 4;485(7397):242–5.

67. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature. Nature Publishing Group; 2012 May 10;485(7397):246–50.

68. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey a. J, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012 Apr 4;485(7397):1–6.

69. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron. Elsevier Inc.; 2012 Apr 26;74(2):285–99.

70. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014 Nov 13;515(7526):216–21.

71. Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS, et al. Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. PLoS Genet. 2012 Jan;8(4):e1002635.

72. Yuen RKC, Thiruvahindrapuram B, Merico D, Walker S, Tammimies K, Hoang N, et al. Whole-genome sequencing of quartet families with autism spectrum disorder. Nat Med. 2015 Feb;21(2):185–91.

73. Lim ET, Raychaudhuri S, Sanders SJ, Stevens C, Sabo A, MacArthur DG, et al. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron. Elsevier Inc.; 2013 Jan 23;77(2):235–42.

74. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. Nature Publishing Group; 2015 Jun 11;47(6):582–8.

75. Yu TW, Chahrour MH, Coulter ME, Jiralerspong S, Okamura-Ikeda K, Ataman B, et al. Using whole-exome sequencing to identify inherited causes of autism. Neuron. Elsevier Inc.; 2013 Jan 23;77(2):259–73.

76. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014 Oct 29;515(7526):209–15.

77. Willsey AJ, State MW. Autism spectrum disorders: from genes to neurobiology. Curr Opin Neurobiol. 2015 Feb;30:92–9.

78. Li J, Cai T, Jiang Y, Chen H, He X, Chen C, et al. Genes with de novo mutations are shared by four neuropsychiatric disorders discovered from NPdenovo database. Mol Psychiatry. 2016 Feb;21(2):290–7.

79. O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science. 2012;338(6114):1619–22.

80. O'Roak BJ, Stessman HA, Boyle EA, Witherspoon KT, Martin B, Lee C, et al. Recurrent de novo mutations implicate novel genes underlying simplex autism risk. Nat Commun. 2014;5:5595.

81. Bernier R, Golzio C, Xiong B, Stessman HA, Coe BP, Penn O, et al. Disruptive CHD8 mutations define a subtype of autism early in development. Cell. 2014 Jul 17;158(2):263–76.

82. Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature. 2009 May 28;459(7246):528–33.

83. Weiss LA, Arking DE, Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism. Nature. 2009 Oct 8;461(7265):802–8.

84. Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, et al. A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet. 2010;19(20):4072–82.

85. Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. Hum Mol Genet. 2012 Nov 1;21(21):4781–92.

86. Chaste P, Klei L, Sanders SJ, Hus V, Murtha MT, Lowe JK, et al. A genome-wide association study of autism using the Simons Simplex Collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? Biol Psychiatry. 2015;77(9):775–84.

87. Berg JM, Geschwind DH. Autism genetics: searching for specificity and convergence. Genome Biol. 2012;13(7):247.

88. Curran S, Bolton P, Rozsnyai K, Chiocchetti A, Klauck SM, Duketis E, et al. No association between a common single nucleotide polymorphism, rs4141463, in the MACROD2 gene and autism spectrum disorder. Am J Med Genet B Neuropsychiatr Genet. 2011 Sep;156B(6):633–9.

89. Devlin B, Melhem N, Roeder K. Do common variants play a role in risk for autism? Evidence and theoretical musings. Brain Res. 2011;1380:78–84.

90. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature. 2014 Jul 24;511(7510):421–7.

91. Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, Willsey a. J, et al. Common genetic variants, acting additively, are a major source of risk for autism. Mol Autism. 2012;3(1):9.

92. Gaugler T, Klei L, Sanders SJ, Bodea C a, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46(8):881–5.

93. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009 Oct 8;461(7265):747–53.

94. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446–50.

95. McCarthy MI, Hirschhorn JN. Genome-wide association studies: potential next steps on a genetic journey. Hum Mol Genet. 2008 Oct 15;17(R2):R156–65.

96. Youssoufian H, Pyeritz RE. Mechanisms and consequences of somatic mosaicism in humans. Nat Rev Genet. 2002;3(10):748–58.

97. Pagnamenta AT, Lise S, Harrison V, Stewart H, Jayawant S, Quaghebeur G, et al. Exome sequencing can detect pathogenic mosaic mutations present at low allele frequencies. J Hum Genet. 2012;57(1):70–2.

98. Biesecker LG, Spinner NB. A genomic view of mosaicism and human disease. Nat Rev Genet. 2013;14(5):307–20.

99. D'Gama AM, Pochareddy S, Li M, Jamuar SS, Reiff RE, Lam A-TN, et al. Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms. Neuron. 2015 Dec 2;88(5):910–7.

100. Lobo I. Multifactorial Inheritance and Genetic Disease. Nat Educ. 2008;1(1):5.

101. Geschwind DH. Autism: Many Genes, Common Pathways? Cell. 2008;135(3):391–5.

102. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. Nat Rev Genet. 2014 Feb;15(2):133–41.

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

104. Pickles A, Starr E, Kazak S, Bolton P, Papanikolaou K, Bailey A, et al. Variable expression of the autism broader phenotype: findings from extended pedigrees. J Child Psychol Psychiatry. 2000;41(4):491–502.

105. Gerdts J a., Bernier R, Dawson G, Estes A. The Broader Autism Phenotype in Simplex and Multiplex Families. J Autism Dev Disord. 2012 Nov 2;43(7):1597–605.

106. Ben-Yizhak N, Yirmiya N, Seidman I, Alon R, Lord C, Sigman M. Pragmatic language and school related linguistic abilities in siblings of children with autism. J Autism Dev Disord. 2011;41(6):750–60.

107. Dawson G, Estes ÆA, Munson ÆJ. Quantitative Assessment of Autism Symptom-related Traits in Probands and Parents : Broader Phenotype Autism Symptom Scale. 2007;523–36.

108. Geschwind DH. Genetics of autism spectrum disorders. Trends Cogn Sci. 2011;15(9):409–16.

109. Robinson EB, St Pourcain B, Anttila V, Kosmicki JA, Bulik-Sullivan B, Grove J, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. Nat Genet. 2016 Mar 21;

110. Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. Nat Genet. 2008;40(6):695–701.

111. Krumm N, O'Roak BJ, Shendure J, Eichler EE. A de novo convergence of autism genetics and molecular neuroscience. Trends Neurosci. 2014 Feb;37(2):95–105.

112. Muhle R, Trentacoste S V, Rapin I. The genetics of autism. Pediatrics. 2004 May;113(5):e472–86.

113. Kim YS, Leventhal BL. Genetic epidemiology and insights into interactive genetic and environmental effects in autism spectrum disorders. Biol

Psychiatry. 2015;77(1):66–74.

114. Lee BK, Magnusson C, Gardner RM, Blomström Å, Newschaffer CJ, Burstyn I, et al. Maternal hospitalization with infection during pregnancy and risk of autism spectrum disorders. Brain Behav Immun. 2015 Feb;44:100–5.

115. Braunschweig D, Duncanson P, Boyce R, Hansen R, Ashwood P, Pessah IN, et al. Behavioral correlates of maternal antibody status among children with autism. J Autism Dev Disord. 2012 Jul;42(7):1435–45.

116. Bauman MD, Iosif A-M, Ashwood P, Braunschweig D, Lee A, Schumann CM, et al. Maternal antibodies from mothers of children with autism alter brain growth and social behavior development in the rhesus monkey. Transl Psychiatry. 2013 Aug 9;3(6146):e278.

117. Hultman CM, Sandin S, Levine SZ, Lichtenstein P, Reichenberg a. Advancing paternal age and risk of autism: new evidence from a population-based study and a meta-analysis of epidemiological studies. Mol Psychiatry. 2011;16(12):1203–12.

118. Sandin S, Hultman CM, Kolevzon A, Gross R, MacCabe JH, Reichenberg A. Advancing maternal age is associated with increasing risk for autism: A review and meta-analysis. J Am Acad Child Adolesc Psychiatry. 2012;51(5):477–86.e1.

119. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-related air pollution, particulate matter, and autism. JAMA psychiatry. 2013;70(1):71–7.

120. Rossignol D a, Genuis SJ, Frye RE. Environmental toxicants and autism spectrum disorders: a systematic review. Transl Psychiatry. 2014;4(2):e360.

121. Wu Ct, Morris JR. Genes, genetics, and epigenetics: a correspondence. Science. 2001;293(5532):1103–5.

122. Trerotola M, Relli V, Simeone P, Alberti S. Epigenetic inheritance and the missing heritability. Hum Genomics. 2015;9(1):17.

123. Hall L, Kelley E. The contribution of epigenetics to understanding genetic factors in autism. Autism. 2014 Nov;18(8):872–81.

124. Melnyk S, Fuchs GJ, Schulz E, Lopez M, Kahler SG, Fussell JJ, et al. Metabolic imbalance associated with methylation dysregulation and oxidative damage in children with autism. J Autism Dev Disord. 2012;42(3):367–77.

125. Ronan JL, Wu W, Crabtree GR. From neural development to cognition: unexpected roles for chromatin. Nat Rev Genet. 2013;14(5):347–59.

126. Parikshak NN, Gandal MJ, Geschwind DH. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat Rev Genet. 2015;16(8):441–58.

127. Purcell a E, Jeon OH, Zimmerman a W, Blue ME, Pevsner J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. Neurology. 2001;57(9):1618–28.

128. Garbett K, Ebert PJ, Mitchell A, Lintas C, Manzi B, Mirnics K, et al. Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiol Dis. 2008;30(3):303–11.

129. Ginsberg MR, Rubin RA, Falcone T, Ting AH, Natowicz MR. Brain transcriptional and epigenetic associations with autism. PLoS One. 2012;7(9):e44736.

130. Chow ML, Pramparo T, Winn ME, Barnes CC, Li H-R, Weiss L, et al. Age-dependent brain gene expression and copy number anomalies in autism suggest distinct pathological processes at young versus mature ages. PLoS Genet. 2012;8(3):e1002592.

131. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Horvath S, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. Nature. 2011;474(7351):380–4.

132. Gupta S, Ellis SE, Ashar FN, Moes A, Bader JS, Zhan J, et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. Nat Commun. 2014;5:5748.

133. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. Nature. 2011;478(7370):483–9.

134. Ben-David E, Shifman S. Networks of neuronal genes affected by common and rare variants in autism spectrum disorders. PLoS Genet. 2012;8(3):e1002556.

135. Parikshak NN, Luo R, Zhang A, Won H, Lowe JK, Chandran V, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. Cell. 2013;155(5):1008–21.

136. Willsey a. J, Sanders SJ, Li M, Dong S, Tebbenkamp AT, Muhle R a., et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2013;155(5):997–1007.

137. Sakai Y, Shaw C a, Dawson BC, Dugas D V, Al-Mohtaseb Z, Hill DE, et al. Protein interactome reveals converging molecular pathways among autism disorders. Sci Transl Med. 2011;3(86):86ra49.

138. Corominas R, Yang X, Lin G, Kang S. Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. Nat Commun. 2014;5:3650.

139. Chang J, Gilman SR, Chiang AH, Sanders SJ, Vitkup D. Genotype to phenotype relationships in autism spectrum disorders. Nat Neurosci. 2015 Feb;18(2):191–8.

140. Guthrie W, Swineford LB, Nottke C, Wetherby AM. Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation. J Child Psychol Psychiatry Allied Discip. 2013;54(5):582–90.

141. Zwaigenbaum L, Bauman ML, Choueiri R, Kasari C, Carter A, Granpeesheh D, et al. Early Intervention for Children With Autism Spectrum Disorder Under 3 Years of Age: Recommendations for Practice and Research. Pediatrics.

2015 Oct 1;136(Supplement):S60–81.

142. Harris SL, Handleman JS. Age and IQ at intake as predictors of placement for young children with autism: a four- to six-year follow-up. J Autism Dev Disord. 2000 Apr;30(2):137–42.

143. Plauche Johnson C, Myers SM, Johnson CP, Myers SM, American Academy of Pediatrics Council on Children With Disabilities. Identification and evaluation of children with autism spectrum disorders. Pediatrics. 2007 Nov;120(5):1183–215.

144. Filipek PA, Accardo PJ, Ashwal S, Baranek GT, Cook EH, Dawson G, et al. Practice parameter: screening and diagnosis of autism: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. Neurology. 2000 Aug 22;55(4):468–79.

145. Myers SM, Johnson CP. Management of Children With Autism Spectrum Disorders. Pediatrics. 2007;120(5):1162–82.

146. Council on Children With Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children With Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. Pediatrics. 2006 Jul;118(1):405–20.

147. Glascoe FP, Sandler H. Value of parents' estimates of children's developmental ages. J Pediatr. 1995 Nov;127(5):831–5.

148. Glascoe FP. Parents' evaluation of developmental status: how well do parents' concerns identify children with behavioral and emotional problems? Clin Pediatr (Phila). 2003 Mar;42(2):133–8.

149. Glascoe FP. Parents' concerns about children's development: prescreening technique or screening test? Pediatrics. 1997 Apr;99(4):522–8.

150. Baron-Cohen S, Wheelwright S, Cox A, Baird G, Charman T, Swettenham J, et al. Early identification of autism by the CHecklist for Autism in Toddlers (CHAT). J R Soc Med. 2000 Oct;93(10):521–5.

151. Robins DL, Fein D, Barton ML, Green JA. The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. J Autism Dev Disord. 2001 Apr;31(2):131–44.

152. Montgomery JM, Duncan CR, Francis GC. Test Review: Siegel, B. (2004). Pervasive Developmental Disorder Screening Test II (PDDST-II). San Antonio, TX: Harcourt. J Psychoeduc Assess. 2007;25:299–306.

153. Carter MT, Scherer SW. Autism spectrum disorder in the genetics clinic: a review. Clin Genet. 2013 May;83(5):399–407.

154. Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. Genet Med. 2013 May;15(5):399–407.

155. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet. The American Society of Human Genetics; 2010 May 14;86(5):749–64.

156. Directors AB of. ACMG Policy statement: points to consider in the clinical application of genomic sequencing ACMG Policy statement. 2012;14(8):759–61.

157. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA. 2014 Nov 12;312(18):1870–9.

158. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. JAMA. 2015 Sep 1;314(9):895–903.

159. Tan R, Wang Y, Kleinstein SE, Liu Y, Zhu X, Guo H, et al. An evaluation of copy number variation detection tools from whole-exome sequencing data. Hum Mutat. 2014 Jul;35(7):899–907.

160. Dorschner MO, Amendola LM, Turner EH, Robertson PD, Shirts BH, Gallego CJ, et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. Am J Hum Genet. 2013;93(4):631–40.

161. Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res. 2015;25(3):305–15.

162. Williams JL, Faucett WA, Smith-Packard B, Wagner M, Williams MS. An assessment of time involved in pre-test case review and counseling for a whole genome sequencing clinical research program. J Genet Couns. 2014 Aug;23(4):516–21.

163. Machini K, Douglas J, Braxton A, Tsipis J, Kramer K. Genetic counselors' views and experiences with the clinical integration of genome sequencing. J Genet Couns. 2014 Aug;23(4):496–505.

164. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013 Jul;15(7):565–74.

165. American College of Medical Genetics and Genomics. Incidental findings in clinical genomics: a clarification. Genet Med. 2013 Aug;15(8):664–6.

166. Directors AB of. Points to consider in the clinical application of genomic sequencing. Genet Med. 2012;14(8):759–61.

167. Turner TN, Hormozdiari F, Duyzend MH, McClymont SA, Hook PW, Iossifov I, et al. Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. Am J Hum Genet. 2016 Jan 7;98(1):58–74.

168. Myers SM, Johnson CP, Plauche Johnson C, Myers SM, Olivié H, Buie T, et al. Clinical practice. Pediatrics. 2007;120(5):741–9.

169. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: a systematic literature review. Eur Child Adolesc Psychiatry. 2013 Jun;22(6):329–40.

170. Lord C, Rutter M, Le Couteur A. 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 Oct;24(5):659–85.

171. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, et al. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord. 1989 Jun;19(2):185–212.

172. McPheeters ML, Warren Z, Sathe N, Bruzek JL, Krishnaswami S, Jerome RN, et al. A systematic review of medical treatments for children with autism spectrum disorders. Pediatrics. 2011 May;127(5):e1312–21.

173. Weitlauf AS, McPheeters ML, Peters B, Sathe N, Travis R, Aiello R, et al. Therapies for Children With Autism Spectrum Disorder. Comp Eff Rev. Agency for Healthcare Research and Quality (US); 2014;(137).

174. Warren Z, McPheeters ML, Sathe N, Foss-Feig JH, Glasser A, Veenstra-Vanderweele J. A systematic review of early intensive intervention for autism spectrum disorders. Pediatrics. 2011 May;127(5):e1303–11.

175. Gray D. Ten years on: A longitudinal study of families of children with autism. J Intellect Dev Disabil. 2002;27(3):215–22.

176. Dumas JE, Wolf LC, Fisman SN, Culligan A. Parenting stress, child behavior problems, and dysphoria in parents of children with autism, down syndrome, behavior disorders, and normal development. Exceptionality. 1991;2(2):97–110.

177. Huddleston J, Eichler EE. An Incomplete Understanding of Human Genetic Variation. Genetics. 2016 Apr;202(4):1251–4.

178. Bailey J a, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet. 2006 Jul;7(7):552–64.

179. Emanuel BS, Shaikh TH. Segmental duplications: an “expanding” role in genomic instability and disease. Nat Rev Genet. 2001;2(10):791–800.

180. Samonte RV, Eichler EE. Segmental duplications and the evolution of the primate genome. Nat Rev Genet. 2002;3(1):65–72.

181. Armengol L, Pujana MA, Cheung J, Scherer SW, Estivill X. Enrichment of segmental duplications in regions of breaks of synteny between the human and mouse genomes suggest their involvement in evolutionary rearrangements. Hum Mol Genet. 2003 Sep 1;12(17):2201–8.

182. Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, Pinkel D,

et al. Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. Genome Res. 2003 Mar;13(3):347–57.

183. Stankiewicz P, Lupski JR. Genome architecture, rearrangements and genomic disorders. Trends Genet. 2002 Feb;18(2):74–82.

184. Mefford HC, Eichler EE. Duplication hotspots, rare genomic disorders, and common disease. Curr Opin Genet Dev. 2009;19(3):196–204.

185. Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, et al. Gene copy number variation spanning 60 million years of human and primate evolution. Genome Res. 2007;17(9):1266–77.

186. Nuttle, Xander et al. Human-specific gene evolution and structural diversity of the chromosome 16p11.2 autism CNV. Presented at American Society of Human Genetics 2014 Annual Meeting San Diego, California. 2014.

187. Zhang L, Lu HHS, Chung W, Yang J, Li W-H. Patterns of segmental duplication in the human genome. Mol Biol Evol. 2005;22(1):135–41.

188. Tuzun E, Sharp AJ, Bailey J a, Kaul R, Morrison VA, Pertz LM, et al. Fine-scale structural variation of the human genome. Nat Genet. 2005;37(7):727–32.

189. Marques-Bonet T, Girirajan S, Eichler EE. The origins and impact of primate segmental duplications. Trends Genet. 2009;25(10):443–54.

190. Lupski JR. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet. 1998 Oct;14(10):417–22.

191. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. Pathogenetics. 2008;1(1):4.

192. Turner DJ, Miretti M, Rajan D, Fiegler H, Carter NP, Blayney ML, et al. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. Nat Genet. 2008 Jan;40(1):90–5.

193. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature. 2006 Nov 23;444(7118):444–54.

194. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, et al. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet. 2005 Jul;77(1):78–88.

195. Estivill X, Cheung J, Pujana MA, Nakabayashi K, Scherer SW, Tsui L-C. Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome. Hum Mol Genet. 2002 Aug 15;11(17):1987–95.

196. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, et al. Linkage disequilibrium and heritability of copy-number polymorphisms

within duplicated regions of the human genome. Am J Hum Genet. 2006 Aug;79(2):275–90.

197. Cooper GM, Zerr T, Kidd JM, Eichler EE, Nickerson DA. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. Nat Genet. 2008 Oct;40(10):1199–203.

198. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, et al. Personalized copy number and segmental duplication maps using next-generation sequencing. Nat Genet. 2009 Oct;41(10):1061–7.

199. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, et al. Diversity of human copy number variation and multicopy genes. Science. 2010 Oct 29;330(6004):641–6.

200. Valero MC, de Luis O, Cruces J, Pérez Jurado LA. Fine-scale comparative mapping of the human 7q11.23 region and the orthologous region on mouse chromosome 5G: the low-copy repeats that flank the Williams-Beuren syndrome deletion arose at breakpoint sites of an evolutionary inversion(s). Genomics. 2000 Oct 1;69(1):1–13.

201. Peoples R, Franke Y, Wang YK, Pérez-Jurado L, Paperna T, Cisco M, et al. A physical map, including a BAC/PAC clone contig, of the Williams-Beuren syndrome--deletion region at 7q11.23. Am J Hum Genet. 2000 Jan;66(1):47–68.

202. Antonell A, de Luis O, Domingo-Roura X, Pérez-Jurado LA. Evolutionary mechanisms shaping the genomic structure of the Williams-Beuren syndrome chromosomal region at human 7q11.23. Genome Res. 2005 Sep;15(9):1179–88.

203. Merla G, Brunetti-Pierri N, Micale L, Fusco C. Copy number variants at Williams-Beuren syndrome 7q11.23 region. Hum Genet. 2010 Jul;128(1):3–26.

204. Antonell A, Del Campo M, Magano LF, Kaufmann L, de la Iglesia JM, Gallastegui F, et al. Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. J Med Genet. 2010 May;47(5):312–20.

205. Morris C a, Mervis CB, Hobart HH, Gregg RG, Bertrand J, Ensing GJ, et al. GTF2I hemizygosity implicated in mental retardation in Williams syndrome: genotype-phenotype analysis of five families with deletions in the Williams syndrome region. Am J Med Genet A. 2003;123A(1):45–59.

206. Sakurai T, Dorr NP, Takahashi N, McInnes LA, Elder GA, Buxbaum JD. Haploinsufficiency of Gtf2i, a gene deleted in Williams Syndrome, leads to increases in social interactions. Autism Res. 2011 Feb;4(1):28–39.

207. Porter MA, Dobson-Stone C, Kwok JBJ, Schofield PR, Beckett W, Tassabehji M. A role for transcription factor GTF2IRD2 in executive function in Williams-Beuren syndrome. PLoS One. 2012;7(10):e47457.

208. Pérez Jurado LA, Wang YK, Peoples R, Coloma A, Cruces J, Francke U. A duplicated gene in the breakpoint regions of the 7q11.23 Williams-Beuren

syndrome deletion encodes the initiator binding protein TFII-I and BAP-135, a phosphorylation target of BTK. Hum Mol Genet. 1998 Mar;7(3):325–34.

209. Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome deletions. Am J Hum Genet. 2003 Jul;73(1):131–51.

210. Howald C, Merla G, Digilio MC, Amenta S, Lyle R, Deutsch S, et al. Two high throughput technologies to detect segmental aneuploidies identify new Williams-Beuren syndrome patients with atypical deletions. J Med Genet. 2006 Mar;43(3):266–73.

211. Dutly F, Schinzel A. Unequal interchromosomal rearrangements may result in elastin gene deletions causing the Williams-Beuren syndrome. Hum Mol Genet. 1996 Dec;5(12):1893–8.

212. Urbán Z, Helms C, Fekete G, Csiszár K, Bonnet D, Munnich A, et al. 7q11.23 deletions in Williams syndrome arise as a consequence of unequal meiotic crossover. Am J Hum Genet. 1996 Oct;59(4):958–62.

213. Somerville MJ, Mervis CB, Young EJ, Seo E-J, del Campo M, Bamforth S, et al. Severe expressive-language delay related to duplication of the Williams-Beuren locus. N Engl J Med. 2005 Oct 20;353(16):1694–701.

214. Van der Aa N, Rooms L, Vandeweyer G, van den Ende J, Reyniers E, Fichera M, et al. Fourteen new cases contribute to the characterization of the 7q11.23 microduplication syndrome. Eur J Med Genet. 52(2-3):94–100.

215. Morris CA, Mervis CB, Paciorkowski AP, Abdul-Rahman O, Dugan SL, Rope AF, et al. 7q11.23 Duplication syndrome: Physical characteristics and natural history. Am J Med Genet A. 2015 Sep 3;

216. Beunders G, van de Kamp JM, Veenhoven RH, van Hagen JM, Nieuwint AWM, Sistermans EA. A triplication of the Williams-Beuren syndrome region in a patient with mental retardation, a severe expressive language delay, behavioural problems and dysmorphisms. J Med Genet. 2010 Apr;47(4):271–5.

217. Osborne LR, Li M, Pober B, Chitayat D, Bodurtha J, Mandel A, et al. A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. Nat Genet. 2001 Nov;29(3):321–5.

218. Hobart HH, Morris CA, Mervis CB, Pani AM, Kistler DJ, Rios CM, et al. Inversion of the Williams syndrome region is a common polymorphism found more frequently in parents of children with Williams syndrome. Am J Med Genet C Semin Med Genet. 2010 May 15;154C(2):220–8.

219. Cuscó I, Corominas R, Bayés M, Flores R, Rivera-Brugués N, Campuzano V, et al. Copy number variation at the 7q11.23 segmental duplications is a susceptibility factor for the Williams-Beuren syndrome deletion. Genome Res. 2008 May;18(5):683–94.

220. Strømme P, Bjørnstad PG, Ramstad K. Prevalence estimation of Williams syndrome. J Child Neurol. 2002 Apr;17(4):269–71.

221. SCHLESINGER BE, BUTLER NR, BLACK JA. Severe type of infantile hypercalcaemia. Br Med J. 1956 Jan 21;1(4959):127–34.

222. WILLIAMS JC, BARRATT-BOYES BG, LOWE JB. Supravalvular aortic stenosis. Circulation. 1961 Dec;24:1311–8.

223. BEUREN AJ, APITZ J, HARMJANZ D. Supravalvular aortic stenosis in association with mental retardation and a certain facial appearance. Circulation. 1962 Dec;26:1235–40.

224. Pober BR. Williams-Beuren syndrome. N Engl J Med. 2010 Jan 21;362(3):239–52.

225. Pober BR, Morris CA. Diagnosis and management of medical problems in adults with Williams-Beuren syndrome. Am J Med Genet C Semin Med Genet. 2007 Aug 15;145C(3):280–90.

226. Wessel A, Gravenhorst V, Buchhorn R, Gosch A, Partsch C-J, Pankau R. Risk of sudden death in the Williams-Beuren syndrome. Am J Med Genet A. 2004;127A(October 2003):234–7.

227. Eronen M, Peippo M, Hiippala A, Raatikka M, Arvio M, Johansson R, et al. Cardiovascular manifestations in 75 patients with Williams syndrome. J Med Genet. 2002 Aug;39(8):554–8.

228. Pober BR, Wang E, Caprio S, Petersen KF, Brandt C, Stanley T, et al. High prevalence of diabetes and pre-diabetes in adults with Williams syndrome. Am J Med Genet C Semin Med Genet. 2010 May 15;154C(2):291–8.

229. Masserini B, Bedeschi MF, Bianchi V, Scuvera G, Beck-Peccoz P, Lalatta F, et al. Prevalence of diabetes and pre-diabetes in a cohort of Italian young adults with Williams syndrome. Am J Med Genet A. 2013 Apr;161A(4):817–21.

230. Stagi S, Bindi G, Neri AS, Lapi E, Losi S, Jenuso R, et al. Thyroid function and morphology in patients affected by Williams syndrome. Clin Endocrinol (Oxf). 2005;63:456–60.

231. Cambiaso P, Orazi C, Digilio MC, Loche S, Capolino R, Tozzi A, et al. Thyroid morphology and subclinical hypothyroidism in children and adolescents with Williams syndrome. J Pediatr. 2007;150(1):62–5.

232. Parfsch CJ, Dreyer G, Gosch A, Winter M, Schneppenheim R, Wessel A, et al. Longitudinal evaluation of growth, puberty, and bone maturation in children with Williams syndrome. J Pediatr. 1999;134:82–9.

233. Morris CA, Demsey SA, Leonard CO, Dilts C, Blackburn BL. Natural history of Williams syndrome: physical characteristics. J Pediatr. 1988 Aug;113(2):318–26.

234. Cherniske EM, Carpenter TO, Klaiman C, Young E, Bregman J, Insogna K, et al. Multisystem study of 20 older adults with Williams syndrome. Am J Med Genet A. 2004 Dec 15;131(3):255–64.

235. Giannotti A, Tiberio G, Castro M, Virgilii F, Colistro F, Ferretti F, et al. Coeliac

disease in Williams syndrome. J Med Genet. 2001 Nov;38(11):767–8.

236. Morris CA, Leonard CO, Dilts C, Demsey SA. Adults with Williams syndrome. Am J Med Genet Suppl. 1990;6:102–7.

237. Morris CA, Mervis CB. Williams syndrome and related disorders. Annu Rev Genomics Hum Genet. 2000;1:461–84.

238. Pankau R, Partsch CJ, Winter M, Gosch A, Wessel A. Incidence and spectrum of renal abnormalities in Williams-Beuren syndrome. Am J Med Genet. 1996 May 3;63(1):301–4.

239. Pober BR, Lacro R V, Rice C, Mandell V, Teele RL. Renal findings in 40 individuals with Williams syndrome. Am J Med Genet. 1993 May 15;46(3):271–4.

240. Sammour ZM, Gomes CM, de Bessa J, Pinheiro MS, Kim CAE, Hisano M, et al. Congenital genitourinary abnormalities in children with Williams-Beuren syndrome. J Pediatr Urol. 2014 Oct;10(5):804–9.

241. Barozzi S, Soi D, Spreafico E, Borghi A, Comiotto E, Gagliardi C, et al. Audiological follow-up of 24 patients affected by Williams syndrome. Eur J Med Genet. 2013 Sep;56(9):490–6.

242. Bedeschi MF, Bianchi V, Colli AM, Natacci F, Cereda A, Milani D, et al. Clinical follow-up of young adults affected by Williams syndrome: experience of 45 Italian patients. Am J Med Genet A. 2011 Feb;155A(2):353–9.

243. Amenta S, Sofocleous C, Kolialexi A, Thomaidis L, Giouroukos S, Karavitakis E, et al. Clinical manifestations and molecular investigation of 50 patients with Williams syndrome in the Greek population. Pediatr Res. 2005 Jun;57(6):789–95.

244. Winter M, Pankau R, Amm M, Gosch A, Wessel A. The spectrum of ocular features in the Williams-Beuren syndrome. Clin Genet. 1996 Jan;49(1):28–31.

245. Greenberg F, Lewis RA. The Williams syndrome. Spectrum and significance of ocular features. Ophthalmology. 1988 Dec;95(12):1608–12.

246. Mervis CB, Klein-Tasman BP. Williams syndrome: cognition, personality, and adaptive behavior. Ment Retard Dev Disabil Res Rev. 2000;6(2):148–58.

247. Mervis CB, John AE. Cognitive and behavioral characteristics of children with Williams syndrome: Implications for intervention approaches. Am J Med Genet Part C Semin Med Genet. 2010;154:229–48.

248. Mervis CB, Morris CA, Klein-Tasman BP, Bertrand J, Kwitny S, Appelbaum LG, et al. Attentional characteristics of infants and toddlers with Williams syndrome during triadic interactions. Dev Neuropsychol. 2003;23(1-2):243–68.

249. Klein-Tasman BP, Li-Barber KT, Magargee ET. Honing in on the social phenotype in Williams syndrome using multiple measures and multiple raters. J Autism Dev Disord. 2011 Mar;41(3):341–51.

250. Woodruff-Borden J, Kistler DJ, Henderson DR, Crawford NA, Mervis CB. Longitudinal course of anxiety in children and adolescents with Williams syndrome. Am J Med Genet C Semin Med Genet. 2010 May 15;154C(2):277–90.

251. Leyfer OT, Woodruff-Borden J, Klein-Tasman BP, Fricke JS, Mervis CB. Prevalence of psychiatric disorders in 4 to 16-year-olds with Williams syndrome. Am J Med Genet B Neuropsychiatr Genet. 2006 Sep 5;141B(6):615–22.

252. Lee SH, Ripke S, Neale BM, Faraone S V, Purcell SM, Perlis RH, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45(9):984–94.

253. Dykens EM. Anxiety, fears, and phobias in persons with Williams syndrome. Dev Neuropsychol. 2003;23(1-2):291–316.

254. Rowe ML. Mastery Motivation in Young Children with Williams Syndrome Or Down Syndrome. 2007.

255. Kennedy J, Kaye D, Sadler L. Psychiatric Diagnoses in Patients with Williams Syndrome and Their Families. Jefferson Journal of Psychiatry. 2006.

256. Jones W, Bellugi U, Lai Z, Chiles M, Reilly J, Lincoln A, et al. II. Hypersociability in Williams Syndrome. J Cogn Neurosci. 2000;12 Suppl 1:30–46.

257. Asada K, Itakura S. Social phenotypes of autism spectrum disorders and williams syndrome: similarities and differences. Front Psychol. 2012;3:247.

258. Gosch A, Pankau R. “Autistic” behavior in two children with Williams-Beuren syndrome. Am J Med Genet. 1994 Oct 15;53(1):83–4.

259. Herguner S, Mukaddes NM. Autism and Williams syndrome: a case report. World J Biol Psychiatry. 2006;7(3):186–8.

260. Reiss AL, Feinstein C, Rosenbaum KN, Borengasser-Caruso MA, Goldsmith BM. Autism associated with Williams syndrome. J Pediatr. 1985 Feb;106(2):247–9.

261. August GJ, Realmuto GM. Williams syndrome: serotonin’s association with developmental disabilities. J Autism Dev Disord. 1989 Mar;19(1):137–41.

262. Gillberg C, Rasmussen P. Brief report: four case histories and a literature review of Williams syndrome and autistic behavior. J Autism Dev Disord. 1994 Jun;24(3):381–93.

263. Edelmann L, Prosnitz A, Pardo S, Bhatt J, Cohen N, Lauriat T, et al. An atypical deletion of the Williams-Beuren syndrome interval implicates genes associated with defective visuospatial processing and autism. J Med Genet. 2007 Feb;44(2):136–43.

264. Kilinçaslan A, Tanidir C, Tutkunkardaş MD, Mukaddes NM. Asperger’s disorder and Williams syndrome: a case report. Turk J Pediatr. 53(3):352–5.

265. Tordjman S, Anderson GM, Botbol M, Toutain A, Sarda P, Carlier M, et al. Autistic disorder in patients with Williams-Beuren syndrome: a reconsideration of the Williams-Beuren syndrome phenotype. PLoS One. 2012;7(3):e30778.

266. Tordjman S, Anderson GM, Cohen D, Kermarrec S, Carlier M, Touitou Y, et al. Presence of autism, hyperserotonemia, and severe expressive language impairment in Williams-Beuren syndrome. Mol Autism. 2013;4(1):29.

267. Klein-Tasman BP, Mervis CB, Lord C, Phillips KD. Socio-communicative deficits in young children with Williams syndrome: performance on the Autism Diagnostic Observation Schedule. Child Neuropsychol. 2007 Sep;13(5):444–67.

268. Klein-Tasman BP, Phillips KD, Lord C, Mervis CB, Gallo FJ. Overlap with the autism spectrum in young children with Williams syndrome. J Dev Behav Pediatr. 2009 Aug;30(4):289–99.

269. Lincoln AJ, Searcy YM, Jones W, Lord C. Social interaction behaviors discriminate young children with autism and Williams syndrome. J Am Acad Child Adolesc Psychiatry. 2007 Mar;46(3):323–31.

270. Richards C, Jones C, Groves L, Moss J, Oliver C. Prevalence of autism spectrum disorder phenomenology in genetic disorders: a systematic review and meta-analysis. The lancet Psychiatry. 2015 Oct;2(10):909–16.

271. Depienne C, Heron D, Betancur C, Benyahia B, Trouillard O, Bouteiller D, et al. Autism, language delay and mental retardation in a patient with 7q11 duplication. J Med Genet. 2007 Jul;44(7):452–8.

272. Torniero C, Dalla Bernardina B, Novara F, Cerini R, Bonaglia C, Pramparo T, et al. Dysmorphic features, simplified gyral pattern and 7q11.23 duplication reciprocal to the Williams-Beuren deletion. Eur J Hum Genet. 2008 Aug;16(8):880–7.

273. Malenfant P, Liu X, Hudson ML, Qiao Y, Hrynchak M, Riendeau N, et al. Association of GTF2i in the Williams-Beuren syndrome critical region with autism spectrum disorders. J Autism Dev Disord. 2012 Jul;42(7):1459–69.

274. McGrew SG, Peters BR, Crittendon JA, Veenstra-Vanderweele J. Diagnostic yield of chromosomal microarray analysis in an autism primary care practice: which guidelines to implement? J Autism Dev Disord. 2012 Aug;42(8):1582–91.

275. Mulle JG, Pulver AE, McGrath JA, Wolyniec PS, Dodd AF, Cutler DJ, et al. Reciprocal duplication of the Williams-Beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. Biol Psychiatry. 2014 Mar 1;75(5):371–7.

276. Dixit A, McKee S, Mansour S, Mehta SG, Tanteles GA, Anastasiadou V, et al. 7q11.23 Microduplication: a recognizable phenotype. Clin Genet. 2013 Feb;83(2):155–61.

277. Parrott A, James J, Goldenberg P, Hinton RB, Miller E, Shikany A, et al. Aortopathy in the 7q11.23 microduplication syndrome. Am J Med Genet Part A. 2015;167(2):363–70.

278. Zarate Y a., Lepard T, Sellars E, Kaylor J a., Alfaro MP, Sailey C, et al. Cardiovascular and genitourinary anomalies in patients with duplications within the Williams syndrome critical region: Phenotypic expansion and review of the literature. Am J Med Genet Part A. 2014;164(8):1998–2002.

279. Mervis CB, Klein-Tasman BP, Huffman MJ, Velleman SL, Pitts CH, Henderson DR, et al. Children with 7q11.23 duplication syndrome: Psychological characteristics. Am J Med Genet Part A. 2015;n/a – n/a.

280. Velleman SL, Mervis CB. Children with 7q11.23 Duplication Syndrome: Speech, Language, Cognitive, and Behavioral Characteristics and their Implications for Intervention. Perspect Lang Learn Educ. 2011 Oct 1;18(3):108–16.

281. Tassabehji M. Williams-Beuren syndrome: a challenge for genotype-phenotype correlations. Hum Mol Genet. 2003 Oct 15;12 Spec No:R229–37.

282. Roy AL, Du H, Gregor PD, Novina CD, Martinez E, Roeder RG. Cloning of an inr- and E-box-binding protein, TFII-I, that interacts physically and functionally with USF1. EMBO J. 1997 Dec 1;16(23):7091–104.

283. Collette JC, Chen X-N, Mills DL, Galaburda AM, Reiss AL, Bellugi U, et al. William's syndrome: gene expression is related to parental origin and regional coordinate control. J Hum Genet. 2009 Apr;54(4):193–8.

284. Bayarsaihan D, Dunai J, Greally JM, Kawasaki K, Sumiyama K, Enkhmandakh B, et al. Genomic organization of the genes Gtf2ird1, Gtf2i, and Ncf1 at the mouse chromosome 5 region syntenic to the human chromosome 7q11.23 Williams syndrome critical region. Genomics. 2002 Jan;79(1):137–43.

285. Hinsley TA, Cunliffe P, Tipney HJ, Brass A, Tassabehji M. Comparison of TFII-I gene family members deleted in Williams-Beuren syndrome. Protein Sci. 2004 Oct;13(10):2588–99.

286. Tipney HJ, Hinsley TA, Brass A, Metcalfe K, Donnai D, Tassabehji M. Isolation and characterisation of GTF2IRD2, a novel fusion gene and member of the TFII-I family of transcription factors, deleted in Williams-Beuren syndrome. Eur J Hum Genet. 2004 Jul;12(7):551–60.

287. Palmer SJ, Taylor KM, Santucci N, Widagdo J, Chan Y-KA, Yeo J-L, et al. GTF2IRD2 from the Williams-Beuren critical region encodes a mobile-element-derived fusion protein that antagonizes the action of its related family members. J Cell Sci. 2012 Nov 1;125(Pt 21):5040–50.

288. Thompson PD, Webb M, Beckett W, Hinsley T, Jowitt T, Sharrocks AD, et al. GTF2IRD1 regulates transcription by binding an evolutionarily conserved DNA motif "GUCE". FEBS Lett. 2007 Mar 20;581(6):1233–42.

289. Ohazama A, Sharpe PT. TFII-I gene family during tooth development: candidate genes for tooth anomalies in Williams syndrome. Dev Dyn. 2007 Oct;236(10):2884–8.

290. Ferrero GB, Howald C, Micale L, Biamino E, Augello B, Fusco C, et al. An atypical 7q11.23 deletion in a normal IQ Williams-Beuren syndrome patient.

Eur J Hum Genet. 2010 Jan;18(1):33–8.

291. van Hagen JM, van der Geest JN, van der Giessen RS, Lagers-van Haselen GC, Eussen HJFMM, Gille JJP, et al. Contribution of CYLN2 and GTF2IRD1 to neurological and cognitive symptoms in Williams Syndrome. Neurobiol Dis. 2007 Apr;26(1):112–24.

292. Dai L, Bellugi U, Chen X-N, Pulst-Korenberg AM, Järvinen-Pasley A, Tirosh-Wagner T, et al. Is it Williams syndrome? GTF2IRD1 implicated in visual-spatial construction and GTF2I in sociability revealed by high resolution arrays. Am J Med Genet A. 2009 Mar;149A(3):302–14.

293. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015 Sep 23;87(6):1215–33.

294. Lucena J, Pezzi S, Aso E, Valero MC, Carreiro C, Dubus P, et al. Essential role of the N-terminal region of TFII-I in viability and behavior. BMC Med Genet. 2010;11:61.

295. National Society of Genetic Counselors' Definition Task Force, Resta R, Biesecker BB, Bennett RL, Blum S, Hahn SE, et al. A new definition of Genetic Counseling: National Society of Genetic Counselors' Task Force report. J Genet Couns. 2006 Apr;15(2):77–83.

296. Biesecker BB. Goals of genetic counseling. Clin Genet. 2001 Nov;60(5):323–30.

297. Hsia YE. 11. The genetic counselor as information giver. Birth Defects Orig Artic Ser. 1979;15(2):169–86.

298. Kessler S. Psychological Aspects of Genetic Counseling. IX. Teaching and Counseling. J Genet Couns. 1997 Sep;6(3):287–95.

299. Meiser B, Irle J, Lobb E, Barlow-Stewart K. Assessment of the content and process of genetic counseling: a critical review of empirical studies. J Genet Couns. 2008 Oct;17(5):434–51.

300. Biesecker BB, Peters KF. Process studies in genetic counseling: peering into the black box. Am J Med Genet. 2001;106(3):191–8.

301. Weil J. Psychosocial genetic counseling in the post-nondirective era: a point of view. J Genet Couns. 2003 Jun;12(3):199–211.

302. Kessler S. Psychological aspects of genetic counseling. XI. Nondirectiveness revisited. Am J Med Genet. 1997 Oct 17;72(2):164–71.

303. Marteau TM, Saidi G, Goodburn S, Lawton J, Michie S, Bobrow M. Numbers or words? A randomized controlled trial of presenting screen negative results to pregnant women. Prenat Diagn. 2000 Sep;20(9):714–8.

304. Welkenhuysen M, Evers-Kiebooms G, D'Ydewalle G. The language of uncertainty in genetic risk communication: framing and verbal versus numerical information. Patient Educ Couns. 2001 May;43(2):179–87.

305. Gates EA. Communicating risk in prenatal genetic testing. J Midwifery Womens Health. 49(3):220–7.

306. Epstein RM, Alper BS, Quill TE. Communicating evidence for participatory decision making. JAMA. 2004 May 19;291(19):2359–66.

307. Henneman L, Marteau TM, Timmermans DRM. Clinical geneticists' and genetic counselors' views on the communication of genetic risks: a qualitative study. Patient Educ Couns. 2008 Oct;73(1):42–9.

308. Ritvo ER, Freeman BJ, Pingree C, Mason-Brothers A, Jorde L, Jenson WR, et al. The UCLA-University of Utah epidemiologic survey of autism: prevalence. Am J Psychiatry. 1989 Feb;146(2):194–9.

309. Chakrabarti S, Fombonne E. Pervasive developmental disorders in preschool children. JAMA. 2001 Jun 27;285(24):3093–9.

310. Icasiano F, Hewson P, Machet P, Cooper C, Marshall A. Childhood autism spectrum disorder in the Barwon region: a community based study. J Paediatr Child Health. 2004 Dec;40(12):696–701.

311. Lauritsen MB, Pedersen CB, Mortensen PB. Effects of familial risk factors and place of birth on the risk of autism: a nationwide register-based study. J Child Psychol Psychiatry. 2005 Sep;46(9):963–71.

312. Grønborg TK, Schendel DE, Parner ET. Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: a population-based cohort study. JAMA Pediatr. 2013 Oct;167(10):947–53.

313. Risch N, Hoffmann TJ, Anderson M, Croen LA, Grether JK, Windham GC. Familial recurrence of autism spectrum disorder: evaluating genetic and environmental contributions. Am J Psychiatry. 2014 Nov 1;171(11):1206–13.

314. Wood CL, Warnell F, Johnson M, Hames A, Pearce MS, McConachie H, et al. Evidence for ASD recurrence rates and reproductive stoppage from large UK ASD research family databases. Autism Res. 2015 Feb 1;8(1):73–81.

315. Parr J. Autism. BMJ Clin Evid. 2010;2010.

316. Grønborg TK, Hansen SN, Nielsen S V, Skytthe A, Parner ET. Stoppage in Autism Spectrum Disorders. J Autism Dev Disord. 2015 Nov;45(11):3509–19.

317. Hoffmann TJ, Windham GC, Anderson M, Croen LA, Grether JK, Risch N. Evidence of reproductive stoppage in families with autism spectrum disorder: a large, population-based cohort study. JAMA psychiatry. 2014 Aug;71(8):943–51.

318. Austin JC, Palmer CGS, Rosen-Sheidley B, Veach PM, Gettig E, Peay HL. Psychiatric disorders in clinical genetics II: Individualizing recurrence risks. J Genet Couns. 2008 Feb;17(1):18–29.

319. Mercer L, Creighton S, Holden JJ a, Lewis MES. Parental perspectives on the causes of an autism spectrum disorder in their children. J Genet Couns. 2006 Feb;15(1):41–50.

320. Selkirk CG, McCarthy Veach P, Lian F, Schimmenti L, LeRoy BS. Parents' perceptions of autism spectrum disorder etiology and recurrence risk and effects of their perceptions on family planning: Recommendations for genetic counselors. J Genet Couns. 2009 Oct;18(5):507–19.

321. Whitelaw C, Flett P, Amor DJ. Recurrence risk in Autism Spectrum Disorder: a study of parental knowledge. J Paediatr Child Health. 2007 Nov;43(11):752–4.

322. Chen LS, Li C, Wang CH, Amuta a, Li M, Huang TY, et al. Autism spectrum disorders: perceptions of genetic etiology and recurrence risk among Taiwanese parents of affected children. Clin Genet. 2014 Sep 30;1–6.

323. Chen L-S, Xu L, Huang T-Y, Dhar SU. Autism genetic testing: a qualitative study of awareness, attitudes, and experiences among parents of children with autism spectrum disorders. Genet Med. 2013 Apr;15(4):274–81.

324. Harrington JW, Patrick P a, Edwards KS, Brand D a. Parental beliefs about autism: implications for the treating physician. Autism. 2006 Sep;10(5):452–62.

325. Shea L, Newschaffer CJ, Xie M, Myers SM, Mandell DS. Genetic testing and genetic counseling among Medicaid-enrolled children with autism spectrum disorder in 2001 and 2007. Hum Genet. 2014 Jan;133(1):111–6.

326. Vande Wydeven K, Kwan A, Hardan AY, Bernstein J a. Underutilization of genetics services for autism: the importance of parental awareness and provider recommendation. J Genet Couns. 2012 Dec;21(6):803–13.

327. Lovell B, Wetherell MA. The psychophysiological impact of childhood autism spectrum disorder on siblings. Res Dev Disabil. 49-50:226–34.

328. Giovagnoli G, Postorino V, Fatta LM, Sanges V, De Peppo L, Vassena L, et al. Behavioral and emotional profile and parental stress in preschool children with autism spectrum disorder. Res Dev Disabil. 45-46:411–21.

329. Sim A, Cordier R, Vaz S, Netto J, Falkmer T. Factors associated with negative co-parenting experiences in families of a child with autism spectrum disorder. Dev Neurorehabil. 2015 Aug 27;1–9.

330. Goepfert E, Mulé C, von Hahn E, Visco Z, Siegel M. Family System Interventions for Families of Children with Autism Spectrum Disorder. Child Adolesc Psychiatr Clin N Am. 2015 Jul;24(3):571–83.

331. Punshon C, Skirrow P, Murphy G. The not guilty verdict: psychological reactions to a diagnosis of Asperger syndrome in adulthood. Autism. 2009 May;13(3):265–83.

332. Spillers JLH, Sensui LM, Linton KF. Concerns about identity and services among people with autism and Asperger's regarding DSM-5 changes. J Soc Work Disabil Rehabil. 2014;13(3):247–60.

333. Giles DC. "DSM-V is taking away our identity": the reaction of the online community to the proposed changes in the diagnosis of Asperger's disorder. Health (London). 2014 Mar;18(2):179–95.

334. Klein S, Sharifi-Hannauer P, Martinez-Agosto JA. Macrocephaly as a clinical indicator of genetic subtypes in autism. Autism Res. 2013 Feb;6(1):51–6.

335. Butler MG, Dasouki MJ, Zhou X-P, Talebizadeh Z, Brown M, Takahashi TN, et al. Subset of individuals with autism spectrum disorders and extreme macrocephaly associated with germline PTEN tumour suppressor gene mutations. J Med Genet. 2005 Apr;42(4):318–21.

336. Tan M-H, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime cancer risks in individuals with germline PTEN mutations. Clin Cancer Res. 2012 Jan 15;18(2):400–7.

337. Spinelli L, Black FM, Berg JN, Eickholt BJ, Leslie NR. Functionally distinct groups of inherited PTEN mutations in autism and tumour syndromes. J Med Genet. 2015 Feb;52(2):128–34.

338. Rodríguez-Escudero I, Oliver MD, Andrés-Pons A, Molina M, Cid VJ, Pulido R. A comprehensive functional analysis of PTEN mutations: implications in tumor- and autism-related syndromes. Hum Mol Genet. 2011 Nov 1;20(21):4132–42.

339. Hendriks YMC, Verhallen JTCM, van der Smagt JJ, Kant SG, Hilhorst Y, Hoefsloot L, et al. Bannayan-Riley-Ruvalcaba syndrome: further delineation of the phenotype and management of PTEN mutation-positive cases. Fam Cancer. 2003;2(2):79–85.

340. Marsh DJ, Kum JB, Lunetta KL, Bennett MJ, Gorlin RJ, Ahmed SF, et al. PTEN mutation spectrum and genotype-phenotype correlations in Bannayan-Riley-Ruvalcaba syndrome suggest a single entity with Cowden syndrome. Hum Mol Genet. 1999 Aug;8(8):1461–72.

341. Eng C. PTEN: one gene, many syndromes. Hum Mutat. 2003 Sep;22(3):183–98.

342. Marcos J, Renau N, Valverde O, Aznar-Laín G, Gracia-Rubio I, Gonzalez-Sepulveda M, et al. Targeting tryptophan and tyrosine metabolism by liquid chromatography tandem mass spectrometry. J Chromatogr A. 2016 Feb 19;1434:91–101.

343. Brunner HG, Nelen M, Breakefield XO, Ropers HH, van Oost BA. Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science. 1993 Oct 22;262(5133):578–80.

344. Piton A, Poquet H, Redin C, Masurel A, Lauer J, Muller J, et al. 20 ans après: a second mutation in MAOA identified by targeted high-throughput sequencing in a family with altered behavior and cognition. Eur J Hum Genet. 2014 Jun;22(6):776–83.

345. Palmer EE, Leffler M, Rogers C, Shaw M, Carroll R, Earl J, et al. New insights into Brunner syndrome and potential for targeted therapy. Clin Genet. 2016 Jan;89(1):120–7.

346. Fehr S, Wilson M, Downs J, Williams S, Murgia A, Sartori S, et al. The CDKL5 disorder is an independent clinical entity associated with early-onset encephalopathy. Eur J Hum Genet. 2013 Mar;21(3):266–73.

347. Rauch A, Wieczorek D, Graf E, Wieland T, Endele S, Schwarzmayr T, et al. Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. Lancet. 2012 Nov 10;380(9854):1674–82.

348. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015 Mar 12;519(7542):223–8.

349. Poot M, Beyer V, Schwaab I, Damatova N, Van't Slot R, Prothero J, et al. Disruption of CNTNAP2 and additional structural genome changes in a boy with speech delay and autism spectrum disorder. Neurogenetics. 2010 Feb;11(1):81–9.

350. Sheng M, Kim E. The Shank family of scaffold proteins. J Cell Sci. 2000 Jun;113 (Pt 1:1851–6.

351. Phelan K, Rogers RC. Phelan-McDermid Syndrome. GeneReviews(®). 1993.

352. Sarasua SM, Dwivedi A, Boccuto L, Rollins JD, Chen C-F, Rogers RC, et al. Association between deletion size and important phenotypes expands the genomic region of interest in Phelan-McDermid syndrome (22q13 deletion syndrome). J Med Genet. 2011 Nov;48(11):761–6.

353. Leblond CS, Nava C, Polge A, Gauthier J, Huguet G, Lumbroso S, et al. Meta-analysis of SHANK Mutations in Autism Spectrum Disorders: a gradient of severity in cognitive impairments. PLoS Genet. 2014 Sep;10(9):e1004580.

354. Brandt T, Desai K, Grodberg D, Mehta L, Cohen N, Tryfon A, et al. Complex autism spectrum disorder in a patient with a 17q12 microduplication. Am J Med Genet A. 2012 May;158A(5):1170–7.

355. Cochoy DM, Kolevzon A, Kajiwara Y, Schoen M, Pascual-Lucas M, Lurie S, et al. Phenotypic and functional analysis of SHANK3 stop mutations identified in individuals with ASD and/or ID. Mol Autism. 2015;6:23.

356. Duyzend MH, Nuttle X, Coe BP, Baker C, Nickerson DA, Bernier R, et al. Maternal Modifiers and Parent-of-Origin Bias of the Autism-Associated 16p11.2 CNV. Am J Hum Genet. 2016 Jan 7;98(1):45–57.

357. van Daalen E, Kemner C, Verbeek NE, van der Zwaag B, Dijkhuizen T, Rump P, et al. Social Responsiveness Scale-aided analysis of the clinical impact of copy number variations in autism. Neurogenetics. 2011 Nov;12(4):315–23.

358. Hu J, Liao J, Sathanoori M, Kochmar S, Sebastian J, Yatsenko SA, et al. CNTN6 copy number variations in 14 patients: a possible candidate gene for neurodevelopmental and neuropsychiatric disorders. J Neurodev Disord. 2015;7(1):26.

359. Córdova-Fletes C, Domínguez MG, Delint-Ramirez I, Martínez-Rodríguez HG, Rivas-Estilla AM, Barros-Núñez P, et al. A de novo t(10;19)(q22.3;q13.33) leads to ZMIZ1/PRR12 reciprocal fusion transcripts in a girl with intellectual disability and neuropsychiatric alterations. Neurogenetics. 2015 Oct;16(4):287–98.

360. Tate CM, Lee J-H, Skalnik DG. CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. FEBS J. 2010 Jan;277(1):210–23.

361. Lee S-A, Tsao TT-H, Yang K-C, Lin H, Kuo Y-L, Hsu C-H, et al. Construction and analysis of the protein-protein interaction networks for schizophrenia, bipolar disorder, and major depression. BMC Bioinformatics. 2011;12 Suppl 1:S20.

362. Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee Y-H, et al. Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci U S A. 2015;112(41):E5600–7.

363. Ramanan VK, Shen L, Moore JH, Saykin AJ. Pathway analysis of genomic data: concepts, methods, and prospects for future development. Trends Genet. 2012 Jul;28(7):323–32.

364. Palmer CGS, Boudreault P, Baldwin EE, Fox M, Deignan JL, Kobayashi Y, et al. Deaf genetic testing and psychological well-being in deaf adults. J Genet Couns. 2013 Aug;22(4):492–507.

ANNEX

Homs A, Codina-Solà M, Rodríguez-Santiago B, Villanueva CM, Monk D, Cuscó I, et al. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. Transl Psychiatry. 2016 Jul 12;6(7):e855–e855.