Characterization of Autism Spectrum Disorder caused by alterations in complex genomic regions: molecular and pathophysiological mechanisms

Mar Costa Roger

TESI DOCTORAL UPF / 2021

DIRECTORS DE LA TESI

Dra. Ivon Cuscó Martí

Genetics Medicine Group of the Clinical and Molecular Genetic Area from Hospital Vall Hebrón

Prof. Luis A. Pérez Jurado

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

Universitat Pompeu Fabra Barcelona

"El ojo no ve lo que el cerebro no conoce"

DH. Lawrence

A la meva àvia Pilar,

als meus pares,

i a l'Alejandro

Agraïments

Sembla mentida que hagin passat quatre anys i mig des que vaig començar al laboratori 422... Quatre anys i mig que donen per a conèixer a moltíssima gent, i més tenint en compte totes les etapes que he viscut des del 422 al 720 i a Vall Hebrón passant pel teletreball. Espero no deixar-me a ningú, però no m'ho tingueu en compte que estic segura d'haver perdut neurones pel camí!

Primer de tot, agrair a la Ivon i al Luis la oportunitat que m'han donat de poder realitzar una tesis doctoral en el camp de la Genètica. Ivon, donar-te les gràcies mil i una vegades, sense tu no hagués pogut arribar fins aquí, literalment. Gràcies per no donar-te per vençuda amb mi i continuar sent la meva directora després del trasllat. Aprecio molt tot el que m'has ensenyat, sobre el TEA, el pensament crític i la vida en general. Luis, gràcies pels consells i recomanacions durant aquests anys.

Moltes gràcies a la gent amb la que he compartit tantes i tantes hores al lab 720 (o 422!). Ha passat moltíssima gent pel grup durant el temps que hi he sigut: Paula, Judith, Raquel, Marcos, Carlos, Francesc, Marivi, Eulàlia, Andrés, Montse, Clara, Roser, Marc, Patricia, Rubén, Laura, Dóra, Ania, Marina, Goretti, Eva, i més! Gràcies per la companyia, pels dinars, les tradicions nadalenques, les Beer sessions, el vòlei i alguna que altra festa!

Paula, moltes gràcies per ser la meva confident durant tots aquests anys, has sigut un pilar fonamental sense el qual tot hagués sigut molt més difícil. Hem compartit moltíssimes hores plegades, gràcies per escoltar-me sempre i no parar boja amb la meva personalitat frenètica! Les pràctiques (mano a mano) no haguessin sigut el mateix amb una altra persona, i tampoc les birres, les patates amb carn del Bambú o els sobretaules amb cafès!

Judith, no sabria per on començar... Tot i coincidir més aviat poc dins el laboratori, sé que m'he emportat una amiga de per vida!! Gràcies per tots els

moments viscuts, sobretot fora de la vida laboral, ja sigui a Sant Pol, a Merlin o a Barcelona. Gràcies per escoltar-me sempre i per tots els consells que m'has donat, ets una font d'inspiració!

Raquel, qui m'entén més que tu? Gràcies per fer que la situació viscuda hagi estat més fàcil del que podria haver sigut. Trobaré a faltar (de fet, ja ho trobo a faltar!) les sessions de psicologia conjunta, sempre tens les idees clares i fas que els problemes es minimitzin. A part d'això, moltíssisisimes gràcies per l'ajuda tècnica, ho saps tot!!

Roser, moltes gràcies pel suport que m'has donat en aquesta última etapa, has sigut gairebé com una altra directora de tesi. Gràcies per totes les reunions (presencials i majoritàriament telemàtiques), per tot el que m'has ensenyat i per fer-me de guia aquests mesos.

Marcos i Carlos, sou els meus àngels de la guarda bioinformàtics. No sabeu quantes vegades m'heu ajudat! Us demano disculpes per ser tan pesada, i una mica inepte, a vegades... Moltes gràcies per tot.

A més de tothom del lab, el PRBB m'ha permès conèixer molta gent magnífica, incloent els integrants dels diversos equips de vòlei: Mónica, Eva, Sira, Marcos i Marc. Gràcies per tots els moments viscuts, els partits i també els post-partits! Mónica, gràcies per portar-nos la millor cheerleader del món. Anna, ets llum. Sabeu que podeu comptar amb mi pel que necessiteu! Als bioevos, moltes gràcies per acollir-nos per dinar i per les converses interessants! A l'Albertito, per ser el millor company de festes. A la Marta i la Clàudia, per fer-nos costat en les lluites importants. Serruchos PRBB seguint camins paral·lels des del 2011!! Gràcies Laura, Andreu i Alberto, per fer que els dinars fossin més distrets de tant en tant i també compartir les penes en moments d'estrès.

En la última etapa, gràcies per l'acollida a tothom de Vall Hebron!

I està clar que aquest camí s'hagués fet molt més difícil sense totes aquelles persones de fóra de l'àmbit professional... Nerea, gràcies per ser la millor

amiga que podia tenir, tan dins com fora de casa. Em complementes i fas que la vida sigui millor amb tu al costat. Gemma i Marta, de ben segur que heu hagut d'aguantar-me moments estressants al pis, gràcies per això i per ajudar-me a desconnectar durant el dia a dia. Gràcies Anna i Raquel per les cerveses de reflexió dels divendres i per fer-nos costat en els moment difícils. Una menció especial per l'Anna que m'ha ajudat amb la portada! Eli i Laura, gràcies per donar-me un toc de frescor en les nostres quedades puntuals. Gràcies també a la Júlia, Judit, Arnau i Xènia. Gràcies a totes per fer-me costat durant aquests anys! Sergi, gràcies per ser-hi sempre, fins i tot des de la distància! Immunenes, gràcies per tots els moments que hem viscut juntes. Des de la uni, la biogimcana i les biofarres fins a tots els viatges posteriors... Mayayo, encantada d'haver viscut una vida paral·lela amb tu, per poder compartir les penes i les alegries! Anna, Mirna, Isshak, Júlia i Agnès, sou família.

Alejandro, gràcies per ser el meu suport incondicional des del principi. M'has fet tirar endavant quan ni jo em veia capaç de fer-ho... Gràcies per ser com ets i transmetre el teu positivisme, la teva energia i la teva alegria. Gràcies per escoltar-me i estimar-me cada dia. Gràcies pels plans passats, presents i sobretot futurs! També gràcies a tots els teus (Sílvia, Teo, Paula i Adrián) per acollir-me com si fos una més de la família!

I per últim, moltes gràcies als meus pares, Marcel i Eulàlia, per totes les lliçons de vida que m'han ensenyat, per estar amb mi en els moments bons i també en els dolents, per aguantar-me les tonteries quan em poso insuportable i per creure en mi sempre. Gràcies a la iaia Teresa, fent èmfasi en tots els àpats que m'has estalviat de cuinar. I també, a la meva àvia Pilar, per què cada vegada que venia a veure't em treies un somriure i em feies desconnectar fent tonteries plegades i parlant idiomes inventats. Has sigut un gran suport, encara que no ho hagis sabut. Us estimo molt.

MOLTES GRÀCIES A TOTXS!

Abstract

Autism Spectrum Disorder (ASD) is a group of neurodevelopmental disorders, the etiology of which is unknown in the majority of cases. Little is known about complex genomic regions containing part of the missing heritability. In this thesis, we have assessed three complex genomic regions including the well-known 7q11.23 *locus* and unexplored 8p23.1 and 17q21.31 *loci* bearing common inversions. We have performed genomic and transcriptomic analyses to unravel common pathophysiological mechanisms in ASD patients with several susceptibility variants (Dup7, inv8p23.1 and inv17q21.31) and also in idiopathic patients considering a multiple-hit model. By using these strategies, we have uncovered several pathways and GO categories, including cholesterol biosynthesis, actin cytoskeletal dynamics, glutamate receptor signaling and immune-inflammatory pathways, that could partially contribute to the underlying molecular mechanism of ASD. Lastly, our data supports the implication of previously ASD-related genes and points out the potential role of novel ASD candidate genes such as *NR4A3, SLC12A6* or *DUSP6*.

Resum

El Trastorn de l'Espectre Autista (TEA) és un grup de trastorns del neuro-desenvolupament, l'etiologia dels quals és desconeguda en la majoria dels casos. Se sap poc sobre les regions genòmiques complexes que podrien contenir part de l'heretabilitat perduda. En aquesta tesi, hem estudiat tres regions complexes incloent el conegut 7q11.23 *locus*, i els inexplorats 8p23.1 i 17q21.31 *loci* que contenen inversions. Hem realitzat anàlisis genòmics i transcriptòmics per trobar mecanismes fisiopatogènics comuns en pacients TEA amb diverses variants de susceptibilitat (Dup7, inv8p23.1 i inv17q21.31) i també en pacients idiopàtics considerant un model de *multiple-hit*. Hem descobert vies i categories d'ontologia gènica, com la síntesi de colesterol, la dinàmica del citoesquelet d'actina, la senyalització dels receptors de glutamat i vies immune-inflamatòries, que podrien contribuir parcialment al mecanisme fisiopatològic del TEA. Per últim, les nostres dades donen suport a la implicació de gens prèviament relacions amb TEA i assenyalen el rol potencial de nous gens candidats com *NR4A3, SLC12A6* o *DUSP6*.

Preface

Given the high genetic heterogeneity of Autism spectrum disorder (ASD), the exploration of complex genomic regions is a good strategy to be approached in ASD patients taking advantage of techniques such as exome sequencing, SNP array and transcriptome sequencing.

This thesis expands the knowledge about certain complex genomic regions and its implications in ASD focusing on regions flanked by segmental duplications that predispose to recurrent rearrangements such as duplications, deletions and inversions. In addition, it addresses deregulated functional pathways that could explain common pathophysiological mechanisms in ASD. The thesis is structured as follows.

The introduction gives a broad overview of the main clinical characteristics of ASD patients and the genetic architecture of ASD. It also covers the role of segmental duplications and inversions in complex genomic regions focusing on 8p23.1 and 17q21.31 *loci* as well as provides a general picture about 7q11.23 complex region and associated disorders. The last part focuses on common pathophysiological mechanisms in ASD.

The main body of the thesis is divided in four chapters addressing the corresponding objectives and with a detailed description of methods and results obtained.

The discussion pretends to globally interpret and integrate the results presented in the chapters of the main body of the thesis. At the same time, it aims to contextualize these results in the previous knowledge of the ASD field.

Finally, conclusions are a summary of the main findings derived from the studies performed in this thesis.

Table of contents

Abstract xi
Preface xiii
List of figures xvii
List of tables xix
List of abbreviations xxi

INTRODUCTION 1

1. Autism Spectrum Disorder 3
    1.1. Epidemiology: history and prevalence 3
    1.2. Diagnosis and screening 4
    1.3. Clinical presentations and co-occurring conditions 8
    1.4. Environmental risk factors 10
    1.5. Evaluation, management and treatment of ASD 12
2. Genetics of Autism Spectrum Disorder 15
    2.1. Genetic models of ASD 15
    2.2. Genetic causes of autism 17
3. Susceptibility factors 26
    3.1. Segmental duplications 26
    3.2. Inversions 28
4. The 7q11.23 recurrent rearrangements 35
    4.1. 7q11.23 genomic structure 35
    4.2. Williams-Beuren syndrome 36
    4.3. 7q11.23 microduplication syndrome 40
5. Molecular pathophysiology of ASD 43

OBJECTIVES 47

CHAPTER 1 51

Genetic factors contributing to autism spectrum disorder in Williams-Beuren syndrome

CHAPTER 2..... 65

Transcriptomic characterization of 7q11.23 patient-specific induced pluripotent stem cells (iPSCs) and derivatives

CHAPTER 3..... 119

Transcriptomic analysis unveils consequences of complex rearrangements associated with Autism Spectrum Disorder

CHAPTER 4..... 155

Identification of recurrently mutated genes and commonly deregulated pathways in Autism Spectrum Disorder multiple-hit model

DISCUSSION 177

CONCLUSIONS 197

REFERENCES 201

ANNEX 229

List of figures

INTRODUCTION

Figure 1. Diagnostic approach for ASD patients 12

Figure 2. Contribution of different genetic variants to ASD genetic liability... 16

Figure 3. Schematic representation of identified genetic causes of ASD..... 17

Figure 4. Genome-wide view of human interchromosomal and inter-chromosomal segmental duplications 28

Figure 5. Formation of inversions by NAHR events mediated by SDs..... 29

Figure 6. Computational methods to detect and genotype inversions 31

Figure 7. Methodology followed by *ScoreInvHap* algorithm..... 32

Figure 8. Schematic view of 17q21.31 inversion..... 32

Figure 9. Schematic representation of the 7q11.23 region..... 36

Figure 10. NAHR events generating the two recurrent deletions in WBS 37

Figure 11. Distinctive facial appearance of subjects with WBS 38

Figure 12. Facial features of Dup7 individuals..... 40

Figure 13. Major cellular pathways affected in ASD 44

Figure 14. NETBAG results from the analysis of genes encompassed in rare *de novo* CNVs..... 46

CHAPTER 1

Figure 1. Strategy followed for the identification of second-hit genetic factors..... 56

Figure 2. Schematic representation of the Williams-Beuren syndrome (WBS) locus 58

Figure 3. Overlap between candidate and functionally constrained genes mutated in our cohort and genes altered in WBS..... 59

Figure 4. Gender (male vs female) comparison of the average number of rare variants in WBS-ASD..... 59

CHAPTER 2

Figure 1. Flowchart illustrating the strategy followed for the transcriptomic analysis 72

Figure 2. Gene expression pattern of genes in the 7q11.23 interval in the different *in vitro* models 81

Figure 3. Differential gene expression and enrichment in fibroblasts 83

Figure 4. Differential gene expression and enrichment analysis in iNeus 85

Figure 5. Weighted gene co-expression network analysis (WGCNA) for iPSCs lines 86

Figure 6. WGCNA analyses for NPCs lines 87

CHAPTER 3

Figure 1. Strategy followed for the transcriptomic study 129

Figure 2. Differentially expressed genes in the 17q21.31 inverted region 132

Figure 3. Differentially expressed genes between inverted individuals (H2/H2) and non-inverted (H1/H1) for 17q21.31 region 133

Figure 4. Differentially expressed genes between inverted ASD individuals and all other groups for the inv17q21.31 134

Figure 5. Differentially expressed genes between I/I ASD individuals and all other groups for inv8p32.1 135

CHAPTER 4

Figure 1. Workflow of the meta-analysis of exome data 160

Figure 2. Genes harbouring rare genetic variants 164

DISCUSSION

Figure 1. Summary of the main findings in the 7q11.23 region 183

Figure 2. Outline of the principal results in 8p23.1 and 17q21.31 unexplored regions 186

Figure 3. Common deregulated functional pathways in ASD patients points towards a probable synaptic dysfunction 191

List of tables

INTRODUCTION

Table 1. DSM-5 Diagnostic criteria for Autism Spectrum Disorder 6

Table 2. Recurrent CNVs strongly associated with ASD 21

Table 3. ASD risk genes identified with TADA model 23

Table 4. Comparison of clinical features observed in 7q11.23 deletion causing WBS and the reciprocal microduplication 42

Table 5. Functions of ASD-associated genes 43

CHAPTER 1

Table 1. Clinical characteristics of the WBS individuals with associated ASD .. 57

Table 2. Summary of the main genetic findings per patient 57

CHAPTER 3

Table 1. TDT meta-analysis of ASD and control trios from multiple datasets for inv17q21.31 and inv8p23.1 131

CHAPTER 4

Table 1. Summarized information of the exome datasets used for the meta-analysis 163

Table 2. Statistically significant deregulated pathways identified in the enrichment analysis 165

DISCUSSION

Table 1. Overview of ASD risk genes as well as novel ASD candidate genes detected in this work 195

List of abbreviations

ABA	Applied behaviour analysis
ADHD	Attention-deficit hyperactivity disorder
ADI-R	Autism Diagnostic Interview-Revised
ADOS-2	Autism Diagnostic Observation Schedule Second Edition
AGP	Autism Genome Project
AMR	Latino/admixed American
APA	American Psychiatric Association
ASD	Autism Spectrum Disorder
BD	Bipolar disorder
BIC	Bayes Information Criteria
BMI	Body mass index
c	centromeric
CAM	Complementary and alternative medicine
CARS-2	Childhood Autism Rating Scale 2
CDC	Centers for Disease Control and Prevention
CMA	Chromosomal microarray
CMT1A	Charcot-Marie-Tooth disease type 1A
CNV	Copy number variant
CPDB	Consensus Path Database
dbGAP	Database of Genotypes And Phenotypes
DEGs	Differentially expressed genes
DP	Depth of coverage
DSM	Diagnostic and Statistical Manual of Mental Disorders
Dup7	7q11.23 microduplication syndrome
EIBI	Early Intensive Behavioral Intervention
ExAC	Exome Aggregation Consortium
FDA	Food and Drug Administration
FDR	False Discovery Rate
FISH	Fluorescence In Situ Hybridization
FXS	Fragile X syndrome
GATK	Genome Analysis Tool Kit
GO	Gene-ontology
GRASP	Genome-wide Repository of Associations between SNPs and phenotypes
GTEx	Genome-Tissue Expression
GWAS	Genome-wide Association studies

HNPP	Hereditary neuropathy with liability to pressure palsies
ICD	International Classification of Diseases
ID	Intellectual disability
IMAGE	International Multi-Center ADHD Genetics Project
iNeus	iPSCs-derived dopaminergic neurons
iPSCs	induced Pluripotent Stem Cells
IQ	Intelligence quotient
JASPER	Joint Attention Symbolic Play Engagement and Regulation
KEGG	Kyoto Encyclopedia of Genes and Genomes
LCR	Low copy repeats
LD	Linkage disequilibrium
LGD	Likely gene-disrupting
LoF	Loss of Function
Log_2FC	logarithm of fold change
lncRNA	Long non-coding RNA
MAF	Minor allele frequency
m	medial
M-CHAT-R/F	Modified Checklist for Autism in Toddlers, Revised, with Follow-up
MCTFR	Minnesota Center for Twin and Family Research
MDS	Multidimensional scaling
MicroRNAs	miRNAs
MMR	Mumps, measles and rubella
NAHR	Non-allelic homologous recombination
NETBAG	Network-Based Analysis of Genes
NFE	non-Finish European
NGS	Next generation sequencing
NHCCs	Non-homologous chromosomal contacts
NIMH	National Institute of Mental Health
NPCs	Neural progenitor cells
OR	Odds ratio
OT	Over-transmission
PCA	Principal component analysis
PCR	Polymerase chain reaction
PECS	Picture Exchange Communication System
PEM	Paired-end sequencing and mapping
PFIDO	Phase Free Inversion Detection Operator
PMM	Postzygotic mosaic mutation

PPD-NOS	Pervasive developmental disorder-not otherwise specified
PPI	Protein-protein interactions
QUAL	Read quality
RA	Rheumatoid arthritis
RNA-seq	RNA sequencing
RR	Relative risk
RRR	Relative recurrence risk
RTT	Rett syndrome
SCQ	Social Communication Questionnaire
SCZ	Schizophrenia
SD	Segmental duplication
SFARI	Simons Foundation Autism Research Initiative
SHARP	SNP Health Asthma Resource Project
SLE	Systemic lupus erythematosus
SNP	Single Nucleotide Polymorphism
SNV	Single nucleotide variant
SRS	Social Responsiveness Scale
SSRI	Selective serotonin reuptake inhibitor
SVAS	Supravalvular aortic stenosis
SWYC	Survey of Wellbeing of Young Children
t	telomeric
TADA	Transmission And De Novo Association
TADs	Topologically associated domains
TDT	Transmission disequilibrium test
TS	Tuberous sclerosis
UMSGARD	University of Miami Study on Genetics of Autism and Related Disorders
VCF	Variant Calling File
WGCNA	Weighted gene co-expression network analysis
WES	Whole exome sequencing
WBS	Williams-Beuren syndrome
WGS	Whole genome sequencing
WHO	World Health Organisation

INTRODUCTION

1. Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental condition characterised by impairments in social-communication skills and the presence of repetitive and restricted behaviours or interests (1).

1.1 Epidemiology: history and prevalence

In 1943, the child psychiatrist Dr. Leo Kanner described 11 children with an "inability to relate themselves in the ordinary way", "desire for maintenance of sameness", good memory and intellectual potential, "limitation in the variety of spontaneous activity", echolalia and sensitivity to loud noises. He coined the term "infantile autism" and was the first to describe this neurodevelopmental disorder (2). Nevertheless, the paediatrician Hans Asperger started in 1938 giving lectures about "Autistic Psychopathy", but it was not until 1944 when he published an article describing 4 patients with impairments in non-verbal communication, avoidance of social contact, lack of eye-to-eye contact, sensory disturbances, clumsy movements and in some cases, special abilities (3). Asperger's work was widely unknown prior to 1981 when Lorna Wing brought it to light (4). Thus, the early recognition of autism is a controversial topic among the scientific community (5).

Estimation of ASD prevalence is a challenging task and differs depending on the methodology and diagnostic criteria used. The first epidemiological study, conducted in 1966, predicted a prevalence of 4.1/10000 for classical autism (6). Since then, prevalence estimates have steadily increased worldwide, probably due to changes in diagnostic criteria, more awareness of the disorder and diagnosis at earlier ages, as well as cultural and environmental factors (7,8). Several reviews have estimated the global ASD prevalence oscillating from 2.67/1000 to 7.6/1000 (9–11). However, prevalence estimates differ greatly among countries, ranging from values of 0.01% in Oman up to 2.7% in South Korea (12,13). The majority of epidemiological studies have been performed in America and Europe. In fact, prevalence is estimated to be ~1.5% in developed

countries (14). In US, the last study from the CDC (Centers for Disease Control and Prevention) estimated an overall ASD prevalence of 1 in 54 (or 1.85%) among children aged 8 years old in 2016 (15,16). In Catalonia, the most recent epidemiological study was based in the administrative data for all children aged 2-17 years insured in the public Catalan Health System during 2017 and found an ASD prevalence of 1.23% (13). It is worth noting that all the above-mentioned studies have been focused on children, but there is one study conducted in England in 2007 that examined the ASD prevalence in adults finding an estimate of 1% (17). All in all, it is important to bear in mind that it is difficult to compare epidemiological studies due to lack of homogeneity in methodological procedures.

Males are more likely to be diagnosed of ASD than females. Despite many studies reporting the ASD male-to-female ratio to be 4:1 (1,16), a recent meta-analysis concluded that the true gender ratio is closer to 3:1 (18). However, in United States the last CDC data found a 4.3:1 male-to-female ratio (15). This review also suggested a possible diagnostic gender bias in which girls meeting criteria for ASD are at higher risk of not receiving a clinical diagnosis. This bias is more noticeable in high-functioning females, whereas girls with comorbid intellectual disability (ID) are more likely to be diagnosed (19). However, even when symptoms in girls and boys are equally severe, girls are less likely to be identified with ASD (20,21). The misrecognition of ASD in females might be a result of gender stereotypes in the diagnostic process as well as a better compensation of girls through a process called "camouflaging", in which they seek coping strategies to overcome socio-communication difficulties such as peer-imitation and masking (22).

1.2 Diagnosis and screening

Diagnosis of ASD can be challenging due to the variation in the degree of symptomatology displayed by affected individuals. Currently, the two most used international classification systems for diagnosing ASD are the 5th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) by the

American Psychiatric Association (APA) and the 11th revision of the International Classification of Diseases (ICD-11) by the WHO (World Health Organization).

In 2013, the APA revised the Diagnostic and Statistical Manual of Mental Disorders publishing the 5th edition (DSM-5) (1). There are two remarkable changes compared to its previous edition, the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR) from 2000 (23).

Firstly, former diagnostic subtypes (Autistic disorder, Asperger disorder, Pervasive developmental disorder-not otherwise specified (PPD-NOS) and Childhood disintegrative disorder) were grouped into a single category in DSM-5 named Autism Spectrum Disorder. Rett's disorder was also considered a subcategory in DSM-IV-TR, but no longer is included as automatically having a diagnosis of ASD according to DSM-5 (1,23). The second relevant change refers to the core symptoms of ASD. In DSM-IV-TR symptoms were divided into three areas: social interaction, social communication and restricted, repetitive and stereotyped patterns of behaviour. In DSM-5, core symptoms have been divided in two areas merging the social interaction and social communication domains into one "social communication" domain, and maintaining the area of restricted and repetitive patterns of activities, behaviours or interests (1,23).

In order to fulfill diagnostic criteria for ASD using DSM-5, all symptoms of social communication area need to be present as well as 2 of 4 features related to restricted and repetitive behaviours. See table 1 (1).

The DSM-5 notes that although symptoms should be present in the early developmental period, they may fully manifest at older ages. Also, the DSM-5 has introduced a rating of severity divided in three levels depending on the support required (1,23). Recent studies have shown that DSM-5 criteria are more stringent, but have good sensitivity and specificity. However, children with previous diagnosis of Asperger's disorder or pervasive developmental disorder and cognitively able individuals are less likely to be diagnosed of ASD using DSM-5 (24,25).

DOMAINS	CRITERIA	EXAMPLES
A. Persistent deficits in social communication and social interaction across multiple contexts	1. Deficits in social-emotional reciprocity	Abnormal social approach, failure of normal back-and-forth conversation, reduced sharing of interests, emotions or affect, failure to initiate or respond to social interactions.
	2. Deficits in nonverbal communicative behaviours used for social interaction	Poorly integrated verbal and nonverbal communication, abnormalities in eye contact or body language, lack of facial expressions.
	3. Deficits in developing, maintaining and understanding relationships	Difficulties adjusting behaviour to suit various social contexts, difficulties sharing imaginative play or making friends, absence of interest in peers.
B. Restricted, repetitive patterns of behaviour, interests or activities	1. Stereotyped or repetitive motor movements, use of objects or speech	Simple motor stereotypies, lining up toys or flipping objects, echolalia, idiosyncratic phrases.
	2. Insistence on sameness, inflexible adherence to routines or ritualized patterns of behaviour	Extreme distress at small changes, difficulties with transitions, rigid thinking patterns, greeting rituals, need to take same route or eat food every day.
	3. Highly restricted, fixated interests that are abnormal in intensity or focus	Strong attachment or preoccupation with unusual objects, excessively circumscribed or perseverative interest.
	4. Hyper- or hyporeactivity to sensory input or unusual interests in sensory aspects of the environment	Apparent indifference to pain/temperature, adverse response to specific sounds or textures, excessive smelling or touching of objects, visual fascination for lights or movement

Table 1. DSM-5 Diagnostic criteria for Autism spectrum disorder. From the Diagnostic and Statistical Manual of Mental Disorders, Fifth edition (copyright 2013) (1). American Psychiatric Association. All rights reserved.

In 2018, the WHO published a new revision of the International Classification of Diseases, the ICD-11, which will come into effect in 2022. In line with DSM-5, previous subcategories from the ICD-10 are grouped under the heading of Autism Spectrum Disorder, characterized by deficits in social communication and by a range of restricted, repetitive, and inflexible patterns of behaviour and interests (26,27).

Although ASD symptoms usually appear in infancy, the average age of diagnosis is at 4-5 years old (28,29). Early recognition of the disorder through surveillance and screening is key for early intervention which can improve the children's outcome, etiologic investigation and reproductive counselling. Surveillance is defined as an ongoing and longitudinal process in which health caregivers collect data to identify children at risk of developmental delays, whereas screening involves the use of standardized tools for specific ages and developmental stages (30).

A systematic surveillance strategy must be provided to the general pediatric population with routine developmental screenings at 9, 18 and 30 months of age as well as autism-specific screening tests at 18 and 24 months according to the American Academy of Pediatrics. Standardized developmental tools should be used when concerns or risks are raised (31). The developmental surveillance process should begin with the collection of the family history in order to identify risk factors including if any member has been diagnosed of ASD, specially a sibling (31). A recent study found that the ASD recurrence rate is sex-specific and it is higher when the older sibling is a female. The probability of ASD recurrence in a younger male sibling is 16.7% if the older sibling is a female and 12.9% when it is a male. Instead, when the younger sibling is a female, the ASD recurrence is 7.6% with an older sister diagnosed of ASD and 4.2% with an older brother (32). Recurrence risk will be discussed in section 2.

Surveillance should also include accurate observations and rigorous records of the child's developmental history in order to evaluate the attainment of age-specific developmental milestones (31). Some of these are considered early symptoms of ASD called "red flags" and they may alert of ASD risk. The American Academy of Neurology and Child Neurology Society suggests the next red flags are indicative for immediate evaluation: no babbling or pointing or other gesture by 12 months, no single words by 16 months, no 2-word spontaneous phrases (not echolalic) by 24 months and any loss of any language or social skills at any age (33).

The most common ASD-specific screening tools for toddlers are the Modified Checklist for Autism in Toddlers, Revised, with Follow-up (M-CHAT-R/F) and Survey of Wellbeing of Young Children (SWYC) (34,35). For school-age children, screening tools available are Social Communication Questionnaire (SCQ) and Social Responsiveness Scale (SRS), among others (36). Screening tests are not diagnostic tools, but they help the identification of children at risk of an ASD diagnosis. However, additional evaluation is required (31).

1.3 Clinical presentations and co-occurring conditions

ASD is considered a complex and clinically heterogeneous condition due to the enormous clinical diversity among subjects. Symptoms usually emerge in early childhood, before 3 years of age and are classified into core and secondary symptoms (1,37). More than 30% of individuals with ASD experience regression of acquired skills happening between 19 and 20 months (38).

As stated before, the core symptoms of ASD are divided into the social-communication area and restricted and repetitive patterns of activities, behaviours or interests (Table 1). In the social-communication domain, individuals show a wide range of symptoms including absence of joint attention, no response to name, not pointing at objects to show interest, lack of social smiling and social interest, impairment understanding other people's feelings and poor eye contact. Individuals with ASD have limited social and imaginative play and are less likely to conceive appropriate peer relationships as well as may prefer solitary activities. Moreover, the majority of affected individuals present language impairments ranging from complete lack of speech to delays in language acquisition. Other common features are echolalia, lack of desire to communicate, inadequate facial expressions and poorly integrated verbal and nonverbal communication (1,29).

Activities, behaviours and interests in ASD individuals are frequently restrictive and repetitive. They tend to have obsessive or intensely focused interests and although the topic of interest can be typical for a child, it is abnormal in intensity. Furthermore, they can be very persistently attached to objects. Stereotyped or repetitive movements such as hand flipping, finger flicking or persistent sniffing can be recognized. Repetitive use of speech and objects such as lining up toys are also frequent symptoms in ASD children. Insistence on sameness, anger at minor changes and ritualistic behaviour can be seen as well. Sometimes, individuals with ASD show hyper- or hyposensitivity to sensory stimuli or sensory aspects of the environment such as fascination for smells or textures or insensitivity to pain (1,29).

Secondary symptoms or co-occurring conditions are present in more than 70% of individuals with ASD and contribute to autism severity. In general terms, co-occurring conditions are classified in developmental, general medical, psychiatric and behavioural categories (39). Some conditions can mask core ASD symptoms delaying its diagnosis (40).

Regarding concurrent developmental conditions, motor problems like hypotonia, motor delays or catatonia are found in approximately 80% of children with ASD and up to 70% present comorbid ID, although estimates may vary if verbal skills are taken into account (39,41). Between 28 and 44% of ASD individuals fulfill criteria for attention-deficit hyperactivity disorder (ADHD). Moreover, tic disorders can be seen, including Tourette syndrome, which is more frequently associated with ASD than expected by chance (39).

Among comorbid general medical conditions, epilepsy, gastrointestinal disturbances, sleep disorders and immune dysregulation can be reported. The risk of epilepsy and seizures is increased in ASD individuals, and a recent review found the median prevalence of epilepsy in people with ASD to be around 12% (42). Brain imaging studies have reported increased total brain volume implicating grey and white matter structures as well as decreased corpus callosum size (43). Gastrointestinal problems are also frequent between children with ASD ranging from 9 to 91% in prevalence reports. Some observed symptoms are abdominal pain, flatulence, diarrhea, reflux and constipation (44). Sleep disorders, mainly insomnia, are reported to affect between 50 and 80% of ASD individuals (39,45).

Psychiatric and behavioural disorders become more apparent in school-aged children and adolescents. Mood disorders are reported to affect up to 70% of ASD individuals, whereas anxiety disorders have been observed in up to 40%. Depression can be seen among 11% of subjects with ASD, which is strongly associated with high rates of suicidal ideation (46,47). Psychotic disorders are mostly reported in adults with ASD ranging from 12 to 17%, but a recent review found an overall pooled prevalence estimate of 4% for schizophrenia spectrum disorders. Oppositional defiant disorder is also frequent among ASD individuals

(16-28%), as well as obsessive-compulsive disorder (9%) and eating disorders (4-5%). Common behavioural problems in children with ASD are aggressiveness, self-injury and pica (39,47).

ASD is a lifelong neurodevelopmental disorder, in which risk of premature mortality has been estimated to increase 2.56-fold compared to matched general population. The markedly higher premature mortality observed in ASD individuals has been associated to a multitude of medical conditions and is particularly increased in females with low-functioning ASD. Suicide rates are reported to be higher in high-functioning ASD individuals (48).

1.4 Environmental risk factors

The etiology of ASD is complex and only recognized in ~25-35% of patients. Although twin studies have enabled to conclude that ASD is highly heritable, it is considered a complex multifactorial disorder with contributions from genetic and environmental factors. Cohort and case-control studies have allowed to shape the possible role of environmental factors in etiology (49). The ASD heritability and genetic factors will be discussed in following sections. Numerous studies have identified many potential environmental risk factors, but only the most relevant will be revised.

Advanced parental age has been involved in ASD risk in many occasions. A recent study that performed an umbrella review concluded that advanced maternal age was associated with increased risk of ASD with convincing evidence, whereas paternal age was graded as highly suggestive evidence (50). Every 10-year increase in maternal and paternal age, the offspring's risk of ASD increased 18 and 21%, respectively (51).

Maternal health factors, including pregnancy-related complications and conditions, have been exhaustively studied. Maternal diabetes, maternal overweight, maternal autoimmune diseases and family history of autoimmune diseases were associated with an increased risk of ASD in offspring (50,52,53). Convincing evidence for association between chronic or gestational

hypertension (odds ratio –OR– 1,48 and 1,37 respectively) and pre-eclampsia (relative risk –RR– 1,32) with ASD was found (50). Moreover, maternal infection during pregnancy is another factor observed to increase ASD risk (52,53).

Medication use during pregnancy is controversial. Despite use of maternal valproate and selective serotonin reuptake inhibitor (SSRI) was related to increased ASD risk, it is important to take into account the presence of confounders such as seizures. No association was found between maternal smoking nor alcohol intake and ASD (50,52). Another polemical topic is exposure to toxins during pregnancy. While the majority of studies detected a positive association between air pollution exposure during pregnancy and ASD, there are some others that show no association. The most detrimental effect on ASD development is found with exposure to particulate matter $PM_{2.5}$ during the third term of pregnancy. Another strong association was observed between pesticide exposure and ASD. For exposure to endocrine-disruptive chemicals and heavy metals, results are inconsistent (54). Nevertheless, breastfeeding and folic acid supplementation during pregnancy were reported to decrease ASD risk (50).

There are several perinatal risk factors that have been linked to neurological vulnerability and increasing risk for ASD. Some examples are birth injury or trauma (RR 4.90), low birth weight (RR 1.63), caesarean section (pooled adjusted OR 1.23) and umbilical cord complications (pooled RR 1.50) (53).

Undoubtedly, the association that has achieved the greatest popular resonance is vaccination, concretely the vaccine for mumps, measles and rubella (MMR). In the late 1990s, a physician from the Royal Free Hospital in London named Andrew Wakefield published an article relating the MMR vaccine with autism. He hypothesized that the measles virus was responsible for inflammatory lesions in the colon that disrupted its permeability letting through neurotoxic proteins that reached the brain causing autism. Although in 2005 Wakefield was accused of research misconduct, conflict of interests and probably falsehood and his article was retracted, the vaccination theory had been quite accepted among the general population. Later, another hypothesis that aroused was that

autism was caused by the exposure to mercury. Some vaccines contained an antiseptic called thimerosal, which was a combination of ethyl-mercury and thiosalicylate (55). At present, there is any scientific evidence proving these hypotheses and no significant association has been found neither for mercury/thimerosal exposure by vaccination nor MMR vaccine (50).

1.5 Evaluation, management and treatment of ASD

When ASD screening result is positive or multiple risk factors have been identified, it is indicated that the child is referred to a multidisciplinary team for a comprehensive ASD evaluation as well as for management and treatment (31). The diagnostic approach followed in ASD patients will be explained in the following sections (Figure 1).

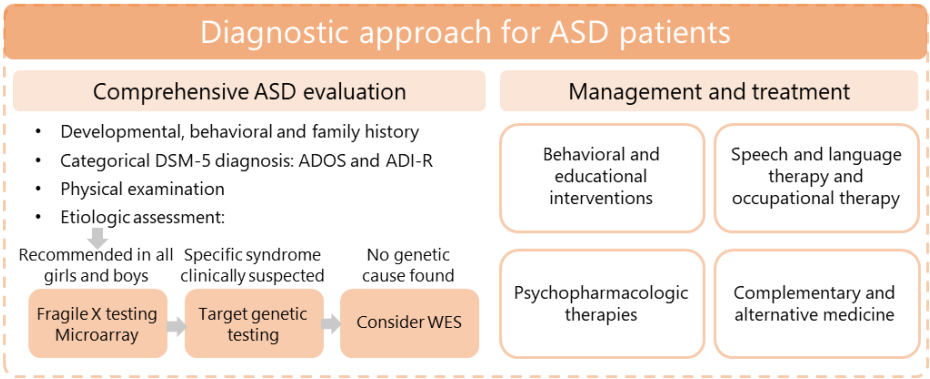


Figure 1. Diagnostic approach for ASD patients. If ASD screening result is positive or risk factors are identified, the diagnostic approach for ASD will start including a comprehensive ASD evaluation and the following management and treatment of the patient.

1.5.1 Comprehensive ASD evaluation

This process should start with a detailed and rigorous assessment of health, developmental and behavioural history including at least a 3-generation family pedigree. It is key to evaluate co-occurring conditions and current functioning of the child as well as to assess cognitive, language and adaptive function, motor and sensory skills and hearing and visual impairment (56).

In order to determine the presence of a categorical DSM-5 diagnosis, a trained specialist may use standardized ASD diagnostic tools such as Autism Diagnostic Observation Schedule Second Edition (ADOS-2) or Autism Diagnostic Interview-

Revised (ADI-R). The ADOS-2 consists of a semi structured standardized evaluation of communication, social interaction and play, whereas the ADI-R is a standardized interview assessing the child's development and current functioning. Other tools used to support the ASD diagnosis are the Childhood Autism Rating Scale 2 (CARS-2) (57).

Another step in the ASD evaluation is the physical examination due to the association of ASD with neurogenetic syndromes (discussed in further sections). It should include assessment of dysmorphic features, growth parameters, skin manifestations, organomegaly and neurologic problems (56,57).

Etiologic assessment is also important in the ASD evaluation process. Although G-banded karyotyping and fragile X testing are commonly performed, chromosomal microarray is considered the first-tier clinical test for ASD patients because it substantially increases the diagnostic yield. However, when there are suspicions for a possible neurogenetic or metabolic syndrome, it is indicated to proceed with appropriate genetic testing. If any of these options succeeds in finding a possible etiology, whole-exome sequencing (WES) should be considered. All families should be offered genetic counselling sessions (56,57).

1.5.2 Management and treatment of ASD

After completing the comprehensive ASD evaluation and concluding with an ASD diagnosis, families should be offered information and support and be guided towards appropriate treatment options. Currently, a cure for ASD does not exist, but treatment strategies are useful for reducing core symptoms and concurrent conditions, improving cognitive ability and adaptive skills and maximizing the child's ability to participate in the community (58).

Treatment strategies should target the child's individual needs, varying with age as well as strengths and weaknesses of every child. Treatment approaches can be divided in four categories, which will be discussed below (57,59).

Behavioural and educational interventions are the mainstay of treatment for ASD and are principally focused on programs for young children including early intensive intervention with low student-to-teacher ratio for at least 25 hours per

week during 12 months. These programs should promote interactions with developing peers and mainly work communication, social and adaptive skills, maladaptive behaviour and cognitive and academic skills. It is recommended to include transition to adult-oriented activities in order to achieve competences needed to promote independence in the workplace and community (59). The gold-standard intervention is Applied behavior analysis (ABA). Other commonly used interventions are Early Intensive Behavioral Intervention (EIBI) and Joint Attention Symbolic Play Engagement and Regulation (JASPER) (57,60).

Other therapeutic interventions include speech and language therapy, which are focused on improving communication skills in ASD children. Gestures, sign language and picture communication programs are commonly used strategies, such as the Picture Exchange Communication System (PECS) (57,59). Occupational therapy is also widely used for ASD treatment in order to improve not only self-care and academic skills, but also adaptive and play skills (56,57).

According to the US Food and Drug Administration (FDA), no medications are approved to treat ASD core symptoms nowadays. However, because of the high prevalence of co-occurring behavioural and psychiatric disorders with ASD, approximately 65% of children diagnosed with ASD receive a psychotropic medication being antipsychotic drugs the most common (57,61). Medication use increases with age, lower levels of adaptive skills and higher challenging behaviours. It is important to bear in mind that adverse effects should be carefully monitored in ASD children as they tend to experience more side effects (57).

The National Center for Complementary and Alternative Medicine (CAM) from the US National Institutes of Health defines CAM as "a group of diverse medical and health care systems, practices and products not presently considered to be part of conventional medicine" (62). They can be divided in natural products (such as dietary modifications), mind and body practices (as acupuncture or yoga) and other therapies like traditional medicine and naturopathy (56). The efficacy of many of these interventions remains unproven, although some have been examined in clinical trials. Evidence has been found supporting dietary supplement of melatonin to treat sleep problems in ASD (56,57).

2. Genetics of Autism Spectrum Disorder

ASD is considered to have a strong genetic component, supported by multiple heritability studies including twin pairs and population-based designs. A recent meta-analysis of twin studies found ASD heritability estimates to range between 64-91% (63). Accordingly, 3 recent population-based studies estimated the heritability of ASD to be approximately 80-85%. Minimal contributions were observed for shared environmental factors and maternal effects described as the association of a maternal phenotype with ASD in offspring (64–66). Hence, the risk for ASD is mostly explained by genetic factors.

The closer the degree of familial relatedness, the higher the risk of ASD. The relative recurrence risk (RRR) is defined as the ASD risk among individuals with an affected family member compared to individuals without any ASD family member. Recurrence risk estimation usually relies on empirical data, which is known as empirical risk, defined as the probability of a trait based on experience instead of based on the inheritance pattern as the causative mechanism is unknown. The ASD RRR for full siblings was reported to be 9.3% (95% CI:8.5-10.1), whereas for maternal and paternal half-siblings was 5.8% (95% CI:4.2-8.1) and 3.9% (95% CI:2.6-5.8), respectively (67). However, in multiplex families with more than one ASD affected child, the RRR was estimated to be around 28.9-50% (68–70).

2.1 Genetic models of ASD

The genetic architecture of ASD is recognized to be complex since different patterns of inheritance have been described including monogenic models and more complex forms such as polygenic and multifactorial models. In addition, ASD genetic liability is shaped by different variants: rare single nucleotide variants (SNVs), copy number variants (CNVs), *de novo* variants (SNVs or CNVs) and common risk alleles from SNPs (Single Nucleotide Polymorphisms). As depicted in figure 2 rare and *de novo* variants account for a moderate or high penetrance, whereas common variants have a small effect size (71,72).

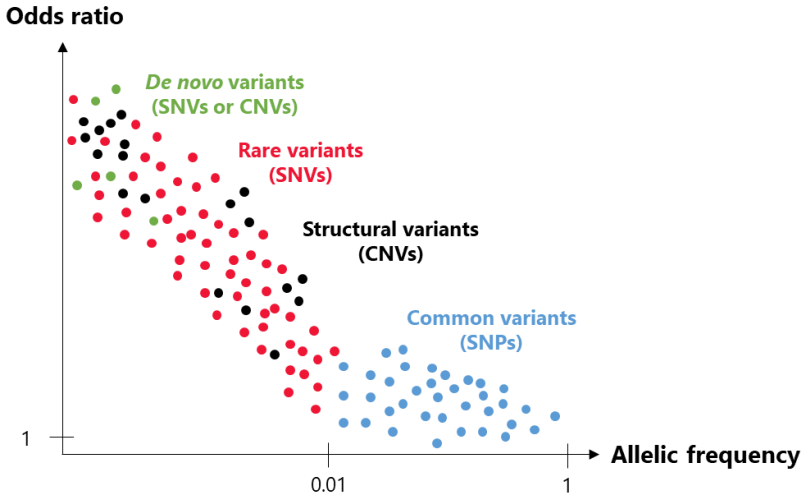


Figure 2. Contribution of different genetic variants to ASD genetic liability. Each dot represents a variant which can contribute to ASD risk: in blue, common risk alleles from SNPs (allelic frequency higher than 0.01 in general population and little effect size); in black, structural variants or copy number variants (CNVs); in red, rare single nucleotide variants (SNVs); and in green, *de novo* mutations which can be either SNVs or structural variants. All these last three categories have an allelic frequency lower than 0.01 in general population and moderate to large effect size. Adapted from (72).

Monogenic models are explained by the rare variant-common disease hypothesis, in which a single highly penetrant gene mutation or CNV is the major contributor to ASD risk (49). Both *de novo* mutations and rare damaging inherited variants can be involved, usually following a Mendelian inheritance, although incomplete penetrance and variable expressivity have been described. The major gene model is supported by the significant higher number of *de novo* mutations found in ASD patients compared to their unaffected siblings (73,74).

Regarding polygenic models, the additive effect of multiple genetic variants accounts for the ASD phenotype. The number and type of mutations involved vary depending on the various models described. The first polygenic model is known as the "two-hits model", in which one rare variant predisposes to the disorder, while a second rare variant is actually the cause. The second variant can also contribute to the phenotypic severity (49). For instance, a recent study found that ASD patients with more than one risk variant were more severely affected regarding cognitive ability than patients carrying one or any risk variants (75).

In the second polygenic model, a rare variant acts additively in a pre-existing genetic background of common risk alleles, whereas the third model assumes that the combination of rare and common variants is the cause of ASD. These first three polygenic models can explain the presence of some *de novo* mutations in healthy population and the fact that some rare variants are inherited from healthy parents (49,72). The common variant-common disease is the fourth polygenic model, in which thousands of common variants (or SNPs), individually of little effect size contribute additively to the overall ASD risk. The sum of these inherited common risk alleles confers a susceptibility above a theoretical risk threshold causing ASD development (49,71).

Lastly, in multifactorial models, it is considered that genotype, environment and *de novo* variation contribute to the genetic landscape of ASD (76).

2.2 Genetic causes of autism

Many studies have focused on unravelling the genetic causes underlying this condition, but ASD etiology is still unclear in many cases known as idiopathic autism. Up to date, only in about 25-35% of patients diagnosed with ASD is possible to recognize the genetic etiology (49). As discussed in the previous section, identified genetic causes can include from a single penetrant mutation causing ASD to a cumulative sum of low-risk alleles, demonstrating the high degree of genetic variability in ASD (Figure 3).

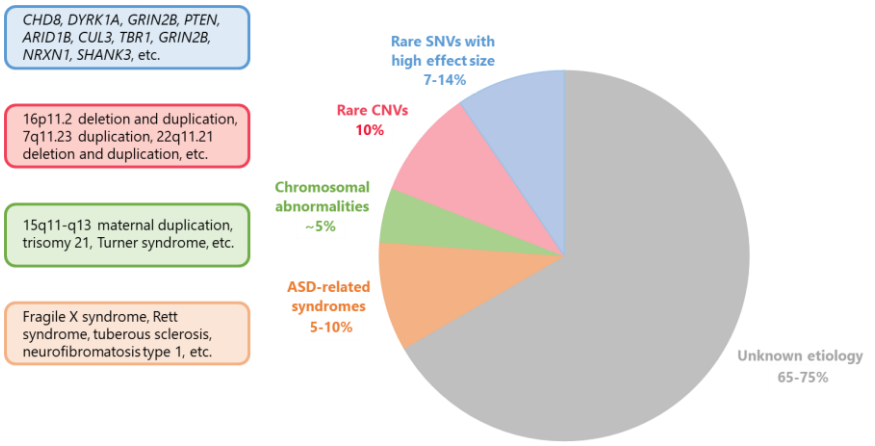


Figure 3. Schematic representation of identified genetic causes of ASD. CNVs: copy number variants, SNVs: single nucleotide variants.

2.2.1 ASD-related syndromes

Among some ASD individuals, autistic features are part of the behavioural manifestation of known genetic syndromes, called syndromic autism. These disorders account for approximately 5-10% of all cases of ASD. Usually, patients with syndromic autism also present dysmorphic features associated to the specific single gene disorder and unlike in idiopathic ASD, male-to-female ratio is equal (77,78).

The most common monogenic disorder related to ASD is fragile X syndrome (FXS), identified in 1.5-3% of individuals with ASD. It is caused by triplet repeat expansions in the X-linked gene *FMR1* (fragile X mental retardation 1), essential for synaptic plasticity. Around 90% of males with full *FMR1* mutations (>200 CGG repeats) display autistic features at some point, and 60% meet ASD diagnostic criteria. In females, up to 23% can be diagnosed of ASD. Macrocephaly with prominent forehead, long face, large protruding ears and prominent chin are typical FXS facial features (49,79).

Mutations in *MECP2* (methyl-CpG-binding protein 2) cause Rett syndrome (RTT), which can be diagnosed in 1-2% of ASD female patients. This X-linked gene is responsible for brain function regulation. Girls affected with RTT develop normally in their first 6 to 18 months until they loss acquired language skills during a regression period and show stereotypic movements (79). Tuberous sclerosis (TS) can be found in approximately 1% of ASD patients and is characterized by hypopigmented macules, brain tumors and multi-organ involvement. Mutations in *TSC1* and *TSC2* (tuberous sclerosis complex 1 and 2) are responsible for TS. These genes are involved in local translation in the central nervous system through rapamycin signaling pathway (mTOR) (49,78). Also, mutations in *PTEN* (Phosphatase and tensin homolog) have been reported in individuals with ASD-macrocephaly syndrome (79).

Other common ASD-related syndromes include neurofibromatosis type 1 (*NF1* gene; ~1%), Williams-Beuren syndrome (WBS; 7q11.23 deletion), 7q11.23 microduplication syndrome, Duchenne muscular dystrophy (*DMD* gene), Timothy syndrome (*CACNA1C* gene), Cohen syndrome (*COH1* gene), Smith-

Magenis syndrome (17p11.2 duplication), DiGeorge syndrome or velocardiofacial syndrome (22q11.2 deletion) and 16p11.2 deletion syndrome. Cornelia de Lange (*SCM1A/SMC3* genes) and CHARGE syndromes can be found in ASD individuals with co-occurring ID and dysmorphic features. Metabolic diseases like phenylketonuria (*PAH* gene) and Smith-Lemli-Opitz syndrome (*DHCR7* gene) as well as some mitochondrial diseases have also been reported (49,78,79).

2.2.2 Chromosomal abnormalities

Cytogenetically visible chromosomal rearrangements are found in 2.2-7.4% of individuals with ASD. These alterations can be detected with karyotyping and fluorescence in situ hybridization (FISH) techniques which have a resolution of >3 Mb and >1 Mb, respectively (77,80). One of the most common chromosomal abnormalities is the 15q11-q13 maternal duplication, observed in 1-3% of ASD patients. Two genes encompassed in this region, *GABRB3* and *UBE3A*, are brain-related genes that codify for GABA receptors and proteasome complex components which have been related to ASD. Individuals with this gain present hypotonia, hypogonadism, fine motor delays, moderate to severe ID, speech and language impairments, epilepsy and other behavioural problems (81).

Other examples of chromosomal rearrangements include 22q11 deletion, 2q37 deletion, 17p11.2 duplication and some aneuploidies such as trisomy 21 (Down syndrome); 45,X Turner syndrome; 47,XXY Klinefelter syndrome and 47,XYY (80,81). Some of these chromosomal abnormalities can also be included in the ASD-related syndromes category.

2.2.3 Copy number variants

The advent of chromosomal microarray (CMA) platforms has enabled the detection of submicroscopic chromosomal abnormalities not detected by conventional karyotype thanks to their higher resolution. Usually, CMA are used to identify CNVs, which are structural variants (duplication and deletion) with a length from 1 Kb to several Mb. CNVs can be either common or rare, transmitted or *de novo*, and have been widely studied in neurodevelopmental disorders (78).

For the last decades, several studies have revealed a higher rate of CNVs in ASD subjects than in controls. Sebat *et al.* found that *de novo* CNVs were significantly associated with ASD as 10% of sporadic ASD patients (families with no previous history of ASD or so-called simplex families) carried a *de novo* CNV, whereas the frequency in multiplex families and healthy controls was 3% and 1%, respectively (82). Further studies encountered similar results with a rate of *de novo* CNVs of 5.8-7.9% in simplex families, 2% in multiplex families and 1.7-2% in siblings (83–85). Moreover, the burden of *de novo* CNVs was higher in affected females than in males (11.7% vs 7.4%) (85). In another study, rare CNVs were significantly more frequent in ASD subjects (4.3%) compared to neurotypical population (2.3%) (86). ASD-specific CNVs were reported in 11.6-12.5% of ASD individuals (87,88). All this suggests that *de novo* CNVs have a major contribution in ASD, but there is also evidence for rare CNVs inherited from apparently normal progenitors that show incomplete penetrance and variable expressivity (89).

The most common ASD-associated CNVs are the microdeletion of 593 Kb in 16p11.2 and the reciprocal microduplication, found in 1% of ASD patients. Interestingly, subjects with the 16p11.2 deletion present macrocephaly whereas duplication carriers have microcephaly (90).

Other recurrent CNVs are 1q21.1 duplication, 7q11.23 duplication, 15q11.2-13 duplication and 22q11.21 deletion and duplication. Some of these CNVs are also involved in other neuropsychiatric disorders like ID, schizophrenia (SCZ) or ADHD. See table 2 for CNVs strongly associated with ASD (91).

Risk Loci	CNV	Gene	Associated neuropsychiatric phenotypes
1q21.1	del/dup		ID, SCZ (both) and ADHD (dup)
2p16.3	del	*NRXN1*	-
3q29	del		ID, speech and language disorder, SCZ, anxiety and bipolar disorder
7q11.23	dup		ID, ADHD, anxiety and oppositional defiant disorder, speech delay
15q11.2-q13.1	dup		ID, ADHD
15q12	dup	*GABRA5*	ID, ADHD, speech delay
15q13.2-q13.3	del/dup		ADHD (del), speech delay (dup) and ID (both)
16p11.2	del/dup		ID (del), SCZ and bipolar disorder (dup)
22q11.21	del/dup		SCZ, ADHD, speech delay and anxiety disorder (del), ID and ADHD (dup)
22q13.33	del	*SHANK3*	ID, language disorder

Table 2. Recurrent CNVs strongly associated with ASD. Other neuropsychiatric phenotypes are also reported. Del: deletion, Dup: duplication, ID: intellectual disability, ADHD: attention deficit hyperactivity disorder, SCZ: schizophrenia. Adapted from (91,92).

In addition, structural variants encompassing ASD susceptibility genes, such as *NRXN1, SHANK3, NLGN1, NLGN4* or *PTCHD1*, have been reported as well (83). Other recurrent ASD-linked loci with a less strong association include 2q37 deletion, 16p13.1 deletion and 17p11.2 duplication (49).

2.2.4 Single nucleotide variants

The advent of next-generation sequencing (NGS) techniques, including WES and whole genome sequencing (WGS), has drastically accelerated gene discovery in ASD highlighting the large number of genes (>1000) that contribute to ASD risk (93,94). In addition, NGS has allowed the identification of SNVs in highly penetrant genes as another cause of ASD. *De novo* SNVs mediating ASD risk have been found in approximately 7-14% of idiopathic ASD cases (91,95,96). Recent analysis concluded that both WES and WGS are efficient diagnostic tools for ASD, especially in idiopathic patients with negative findings in microarrays (75,97).

The first NGS studies focused on simplex families using WES technology in a trio based design to investigate *de novo* events, and in some cases unaffected

siblings were also evaluated (73,95,96,98,99). The first conclusion drawn from these studies was that *de novo* Loss of Function (LoF) or likely gene-disrupting (LGD) mutations, including nonsense, frameshift and splice site variants, were twice as frequent in affected probands than in unaffected siblings (73,95,96). It has been estimated that in probands approximately 41-46% of *de novo* LGD events contribute to ASD risk (73,91). Moreover, *de novo* LoF variants were found at higher rates in affected females than in affected males, and other studies observed similar results with a 2-fold enrichment of *de novo* LGD mutations in females. This fact is in accordance with the female protective effect hypothesis, which states that females require more genetic load to reach the threshold of ASD diagnosis (100,101). Other conclusions extracted from the first studies were that *de novo* mutations were biased towards a paternal origin (4:1) and were positively related to parental age (73,96,98).

The identification of multiple *de novo* variants in the same gene from unrelated individuals pointed out candidate ASD-risk genes, which allowed the development of target sequencing studies of candidate genes (102–104). 1% of sporadic ASD cases were estimated to be explained by *de novo* disruptive mutations in six recurrent genes: *CHD8*, *DYRK1A*, *GRIN2B*, *TBR1*, *PTEN* and *TBL1XR1*. Variants in *CHD8* accounted for 0.35% of ASD cases and were associated with macrocephaly, while patients carrying *DYRK1A* mutations had significantly smaller head sizes (98). A more recent target sequencing study found that severe *de novo* variants including missenses in *SCN2A* and *CHD8* accounted for 1.5% of ASD cases and identified novel macrocephaly-associated ASD genes (*WDFY3, GIGYF2* and *KMT5B*) (104). It is very useful to clinically characterize ASD patients sharing variants in the same gene in order to identify specific subphenotypes.

Subsequent studies not only investigated *de novo* events, but also rare inherited truncating variants using both WES and WGS strategies (100,105–108). Rare inherited truncating SNVs in highly conserved genes were found enriched in probands compared to their unaffected siblings (100,106). A recent study by Satterstrom *et al.*, which is the largest exome sequencing analysis to date,

observed for the first time a 2.1-fold enrichment of *de novo* damaging missense variants in cases compared to controls (100). In line with previous studies, *de novo* variants were mainly originated in the father, whereas rare inherited truncating SNVs showed a bias towards a maternal origin (106,107).

Taking advantage of large sample sizes, several studies carried out a weighted, statistical model known as TADA (Transmission And *De novo* Association) that integrates *de novo*, transmitted and case-control variation to identify ASD risk genes (91,105,109). Following this approach, 33 ASD risk genes were identified with an FDR<0,1 (False Discovery Rate) and 107 genes with an FDR<0,3. These genes encoded proteins implicated in three pathways: chromatin remodeling, transcription and splicing, and synaptic function (105). Another study found 65 ASD risk genes with an FDR≤0,1 involved in chromatin regulation and synapse (91). Recently, 69 genes were significantly associated with ASD (FDR<0,1), from which 16 were newly identified genes (109). All ASD risk genes described in the above-mentioned articles are summarized in table 3.

dnLoF count	FDR ≤ 0.01	0.01 ≤ FDR ≤ 0.05	0.05 ≤ FDR ≤ 0.1
≥2	***CHD8***, ***SCN2A***, ***ARID1B***, ***SYNGAP1***, ***DYRK1A***, CHD2, ***ANK2***, KDM5B, ***ADNP***, ***POGZ***, KMT5B, ***TBR1***, ***GRIN2B***, DSCAM, KMT2C, TCF7L2, TRIP12, ***ASH1L***, ***CUL3***, ***KATNAL2***, GIGYF1, SUV420H1, TNRC6B, WACNCKAP1,	RANBP17, KDM6B, ILF2, SPAST, FOXP1, AKAP9, CMPK2, DDX3X, ASXL3, ***BCL11A***, MLL3, WDFY3, CACNA2D3, PHF2	KMT2E, DIP2A
1	NRXN1, SHANK2, ***PTEN***, SHANK3, ***SETD5***	DNMT3A, ***MYT1L***, RAPGEF4, PRKAR1B, ***GABRB3***, RELN, KAT2B, MFRP, P2RX5	***ETFB***, ***CTTNBP2***, INTS6, USP45, ERBIN, TMEM39B, TSPAN4, MLANA, SMURF1, C16orf13, BTRC, CCSER1, FAM98C, APH1A, CD42BPB, NAA15, MYO9B, NR3C2, TRIO, IRF2BPL, MBD5, NINL, OR52M1, PTK7
0	-	SLC6A1, ZNF559, CAPN12, GRIA1, MIB1	PCM1, MYO5A, UIMC1, VIL1, ACHE, NLGN3

Table 3. ASD risk genes identified with TADA model. Adapted from (91,105,109). Genes in bold are the highly confident subset of genes observed significantly associated with ASD in the three studies. Genes shown underlined are found in two of the articles. If genes were in different categories, prevailed the one with higher FDR.

Common variants, such as SNPs (occurring in more than 1% of the population) could also contribute with small effect size to ASD risk. In fact, the variance in liability accounted for measured SNPs was estimated to be 49.4% (110).

However, a recent meta-analysis found that SNPs-based heritability was 11.8% (111). Anyhow, the contribution of common variants to ASD liability seems to be significant.

Genome-Wide Association studies (GWAS) allow to explore the association between SNPs and human traits like diseases. First GWAS studies investigating ASD failed to achieve genome-wide significant signals (112). The first study to find variants with genome-wide significant association with ASD found a SNP on 5p14.1 (rs4307059-T $p=10^{-10}$) that was related to a long noncoding RNA (lncRNA) named *MSNP1AS* (113). Later on, a meta-analysis identified a single association marker at rs1409313-T (OR=1.12 $p=1.058 \cdot 10^{-8}$) on chromosome 10q14.32 in linkage disequilibrium with a gene-rich region. Among these genes, *CUEDC2* and *PITX3* were related to the ubiquitination-proteasomal degradation pathway and neuronal differentiation, respectively (114). The most recent GWAS based in a case-cohort population with an increased sample size found five common risk variants robustly associated with ASD. These genome-wide significant signals included rs910805 on 20p11.23, rs10099100 on 8p23.1, rs201910565 on 1p21.3, rs71190156 on 20p12.1 and rs111931861 on 7q22.3. Additionally, 7 more loci were identified shared with other traits (schizophrenia, major depression and educational attainment) at equally strict significance levels (111).

2.2.5 Mosaic events

Most of the variants described above are constitutional variants present in every cell (or the majority) in an individual. However, sometimes these variants can be found in mosaicism, which is defined as the presence of two or more genetically different cell populations in the same individual as a result of postzygotic mutations. Depending on the moment after fertilization when the mutation occurs and the cell lineage, they can generate somatic mosaicism or gonadal mosaicism. *De novo* mutations present as heterozygous variants in every cell of an organism are thought to occur in the parental germline, but a considerable part of *de novo* mutations are likely to have arisen postzygotically rather than in parental gametes (115).

Postzygotic mosaic mutations (PMMs) have been implicated in cancer and neurodevelopmental disorders like epilepsy and genetic disorders of neuronal migration (116). However, the detection of PMMs is challenging due to the limited sensitivity to detect this type of mutations and the availability of specific tissues. Although brain would be the tissue of choice for ASD, it is not accessible, so whole blood can be a good approach. In fact, a study carried out deep sequencing in postmortem ASD brains and found two cases with deleterious somatic mutations which might contribute to ASD risk (117). Further studies using WES technology estimated that 4.2-7.5% of all *de novo* mutations were actually PMMs (118–120). In addition, mosaic events were significantly enriched in ASD probands compared to their unaffected siblings as well as ASD probands harbored more deleterious PMMs in brain-expressed critical exons than their siblings (118,119). Mosaic mutations have been identified in previously associated ASD genes such as *SCN2A, SYNGAP1, KMT2C* and *TBL1XR1,* among others (120). Therefore, the study of PMMs is also useful in ASD gene discovery.

3. Susceptibility factors

In complex models (polygenic and multifactorial) explained in previous sections, the sum of multiple genetic variants and environmental factors is responsible for the ASD phenotype, although the number, type of mutations and the environmental factors involved can vary. Most of these variants are susceptibility factors that arouse risk for developing ASD (49). Nowadays, the majority of studies that try to understand the genetic causes of ASD have focused on monogenic causes and/or highly penetrant variants. Therefore, single copy regions and coding regions have been exhaustively explored in ASD patients, but little is known about complex regions and non-coding regions, which may contain part of the lost heritability.

3.1 Segmental duplications

Segmental duplications (SDs) are part of these complex unexplored regions. SDs, also known as low-copy repeats (LCRs), are segments of DNA ranging from 1 to 200 Kb in length with high homology (>95%) found in two or multiple genomic locations. First studies using FISH analysis and computational approaches estimated the human SD content to be around 3.6-5% (121,122). However, a more recent study developed a new computational tool which identified a total of 219 Mbp (7,6%) with SDs in the human genome (123).

SDs can be organised in tandem or interspersed locations around the human genome. Although they have been identified in all human chromosomes, LCRs are non-randomly distributed being enriched in chromosomes 22 and Y. In addition, SDs can be divided into two groups including interchromosomal duplications, in which duplicated segments are found among non-homologous chromosomes and intrachromosomal duplications that are located in a specific chromosome or chromosomal arm (121) (Figure 4). SDs have been observed to cluster together with up to 10-fold enrichment in pericentromeric and subtelomeric regions (122,124). Intrachromosomal SDs are more prone to

contain genes than interchromosomal duplications. Moreover, SDs containing gens tend to be longer and to have higher sequence similarity (124).

Segmental duplications are a common architectural feature of the genomes of primates and are relevant for genetic diversity and evolution. SDs have appeared relatively recently during primate evolution. They have been estimated to emerge during the past 35 million years and have rapidly evolved since the divergence of great apes (125). In addition, SDs can mediate chromosomal structural variation related to genomic disorders. Due to their similarity, they can initiate non-allelic homologous recombination (NAHR) events in which there is a misalignment of highly homologous SDs followed by paralogous recombination. These events can lead to recurrent rearrangements including deletions, duplications, translocations, inversions and marker chromosomes. The type of resulting rearrangement will depend on the orientation of the repeats. In fact, these alterations are found with a 4-12-fold greater frequency close to SDs (126–128).

A study identifying potential hot spots of genomic instability by the presence of SDs mapped 169 large regions flanked by LCRs, from which 24 regions had already been related to genomic disorders (Figure 4) (129). For example, in the chromosome region 17p12, there are two 24 Kb SDs called CMT1A-REP flanking a unique copy region. NAHR between these repeats can result in a 1.5 Mb deletion associated with hereditary neuropathy with liability to pressure palsies (HNPP), whereas the reciprocal microduplication causes Charcot-Marie-Tooth disease type 1A (CMT1A) (126).

Moreover, SDs are important in the evolution of novel genes. One of the mechanisms to create new genes is the duplication of an entire gene due to the presence of SDs followed by positive selection. It is thought that this mechanism is one of the forces driving vertebrates to achieve its proteome diversity and morphological complexity (125,128). Many studies are trying to elucidate the function of these novel genes. Recently, some genes derived from human-specific SDs (*SRGAP2C* and *ARHGAP11B*) have been associated with enlargement of human brain. Therefore, it seems that the advent of novel genes

mediated by SDs and the progress of new function have played a key role in human evolution and adaptation (130).

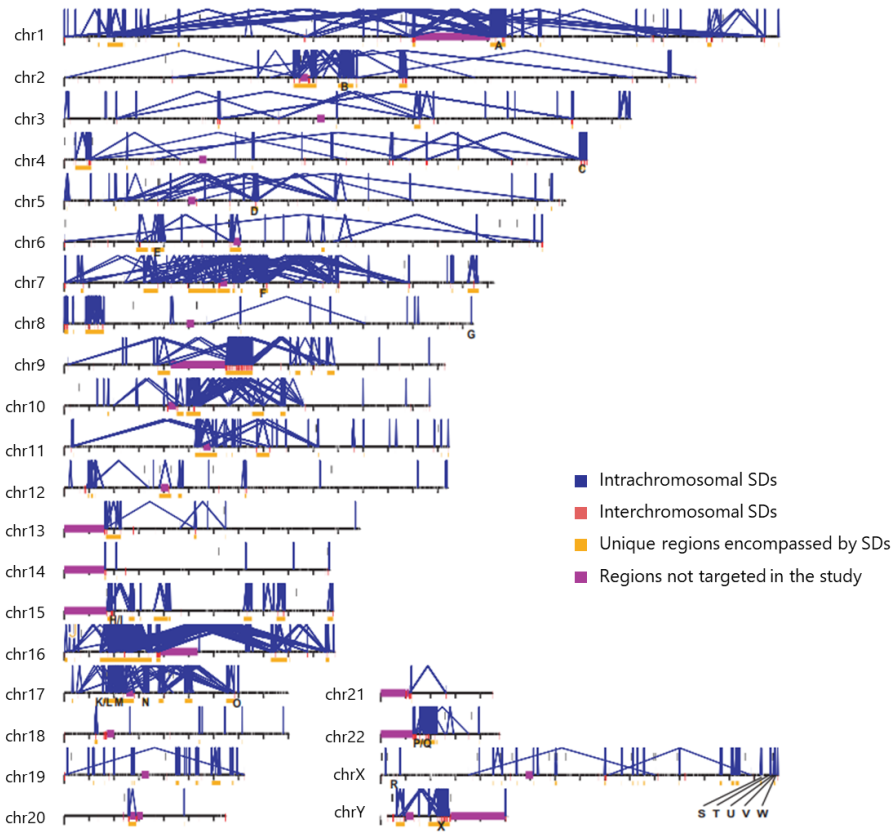


Figure 4. Genome-wide view of human intrachromosomal and interchromosomal segmental duplications. Locus with a letter indicate those regions associated with genomic disorders. SDs: segmental duplications. Adapted from (129).

3.2 Inversions

Inversions are described as a change in orientation of a specific DNA segment. They are classified in two classes including pericentric and paracentric inversions, which include or not the centromere, respectively (131,132). The most common mechanism to originate inversions is NAHR mediated by SDs when these repeats are oppositely oriented on the same chromosome (Figure 5). However, if no high sequence similarity is found at the breakpoints, there are other mechanisms that can generate inversions, such as double-strand break

repair mechanisms or microhomology mediated replication-based mechanisms (132,133).

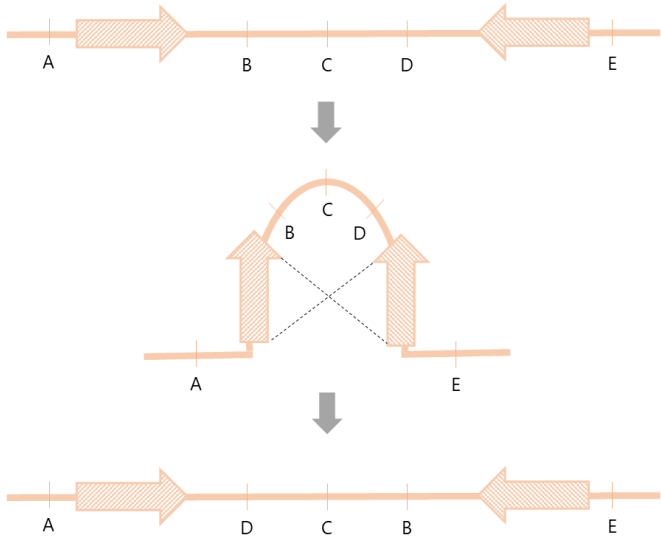


Figure 5. Formation of inversions by NAHR events mediated by SDs. Chromosomes are represented by orange lines, whereas segmental duplications are shown as big arrows with opposite orientation. Capital letters above the line represent unique sequences, which after the NAHR will be inverted in the region flanked by SDs. Adapted from (126,127).

Inversions are known to contribute to chromosomal evolution and are considered a source of genetic variability. These can be explained by indirect and direct effects of inversions. Indirectly, they cause suppression of recombination of the inverted regions in heterozygotes, which over time leads to linkage disequilibrium (LD) between haplotypes (131,132). On the other hand, although inversions are usually considered neutral variants, inversion breakpoints can directly cause gene disruption, alter gene expression due to separation of regulatory elements and in some cases predispose to genetic copy number imbalances (132). For instance, the disruption of *IDS* (Iduorante-2-Sulphatase) due to the presence of an inversion is a common cause of Hunter syndrome (134). Also, the frequency of the 7q11.23 inversion is higher among transmitting parents of Williams-Beuren syndrome (WBS) patients indicating that the inverted allele may predispose to the 7q11.23 deletion (135). Therefore, inversions have been linked to multiple diseases, including being susceptibility factors for different multifactorial pathologies such as autism or obesity (136–138).

3.2.1 Methods for the detection and genotyping of inversions

Initially, inversions were detected by G-banded karyotyping and FISH techniques with limited resolution. Antonacci *et al.* characterize six large inversion polymorphisms in 3 HapMap populations using a metaphase FISH-based assay (139). Other experimental techniques, including PCR (Polymerase Chain Reaction) and inverse PCR, have also been used for targeted studies of inversions. For the discovery of genome-wide inversions, sequencing techniques like WGS comparison and paired-end sequencing and mapping (PEM), have implicated a major breakthrough.

However, these techniques are limited by their relatively low throughput, and for that reason computational algorithms are also useful in the detection of inversions (132). The majority of computational methods focus on the identification of specific genomic patterns caused by inversions.

A powerful tool in the detection of inversions is InveRsion algorithm, which is able to predict inversions from SNP data (140). This tool is useful to find inversions as can predict putative regions harbouring inversions without previous knowledge of breakpoints. inveRsion scans the genome by measuring changes of LD between SNPs blocks across inversion breakpoints. It assumes that close SNP blocks from inversions have higher LD than distant blocks and can detect regions where this pattern has been disrupted, suggesting the presence of a chromosomal inversion. The algorithm quantifies these inversion signals with the Bayes Information Criteria (BIC), and a positive BIC value indicates that the region likely contains an inversion (Figure 6A) (141). However, inveRsion tool was conceived to detect inversion regions, so other algorithms are needed for inversion genotyping.

As explained above, inversions cause suppression of recombination leading to differences between inverted and standard chromosomes. Therefore, it is possible to find variants exclusive from one of the inversion status such as tag SNPs, which are SNPs that individually differentiate the inverted and standard chromosome. The simplest approach for inversion genotyping is the identification of tag SNPs, but unfortunately many inversions do not have tag

SNPs linked (142,143). Clustering methods used for inversion genotyping are an extension of tag SNPs. They are based on a dimensionality reduction technique, such as principal component analysis (PCA) or multidimensional scaling (MDS), on all SNPs harboured in an inversion that allows to cluster haplotype groups due to differences in the mutational content between inverted and standard chromosomes. The first tool implementing this methodology was PFIDO (Phase Free Inversion Detection Operator), which performed a MDS on all the SNPs from the inversion interval in order to cluster inversion haplotypes (144). Later on, InvClust package was developed, which is an algorithm similar to PFIDO that infers inversion status considering the ancestry background, a possible confounder factor in clustering methods (Figure 6B) (145).

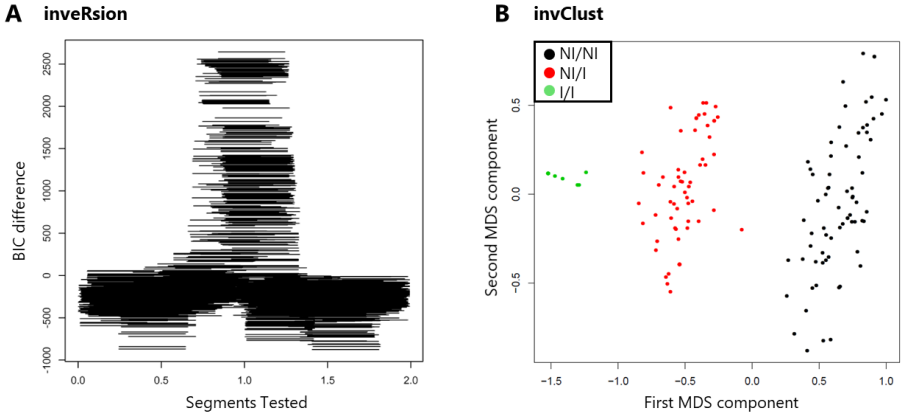


Figure 6. Computational methods to detect and genotype inversions. (A) InveRsion: scan search for inversions showing BIC for each possible window. From (146). (B) InvClust: 17q21.31 inversion genotyping in own data. BIC: Bayes information criteria, MDS: multidimensional scaling, NI: non-inverted, I: inverted.

Nevertheless, these methods face important limitations such as the need of a minimum number of individuals to accurately compute calls, the inefficiency in genotyping large cohorts and problems with inversion calling harmonization in meta-analysis. A newer approach that overcomes these limitations is scoreInvHap, which can reliably genotype 20 human inversions from SNP data. This tool uses a set of experimentally validated reference haplotype-genotypes from the inversion of interest and classifies each individual according to their SNP similarities to the reference genotypes (Figure 7) (147).

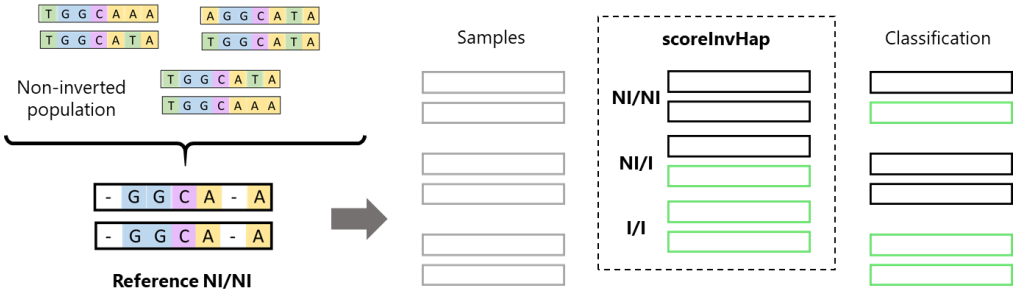


Figure 7. Methodology followed by ScoreInvHap algorithm. Validated haplotype-genotype references are obtained for a following comparison of SNP similarities with samples of interest. NI: non-inverted, I: inverted.

3.2.2 Inversion 17q21.31

The 17q21.31 inversion is a ~970 Kb inversion located in chr17: 43,544,138-44,633,937 (hg19) with two LCRs flanking its breakpoints (Figure 8). It is one of the most structurally complex and evolutionarily dynamic regions of the genome (139,148). There are two highly divergent haplotype groups, H1 (direct orientation) and H2 (inverted) (149). The two status can be distinguished by genotyping a marker (*DG17S142*) in intron 9 of *MAPT* (Microtubule-Associated Protein Tau) and a characteristic 238 bp deletion in the same intron of H2 haplotype (139,148).

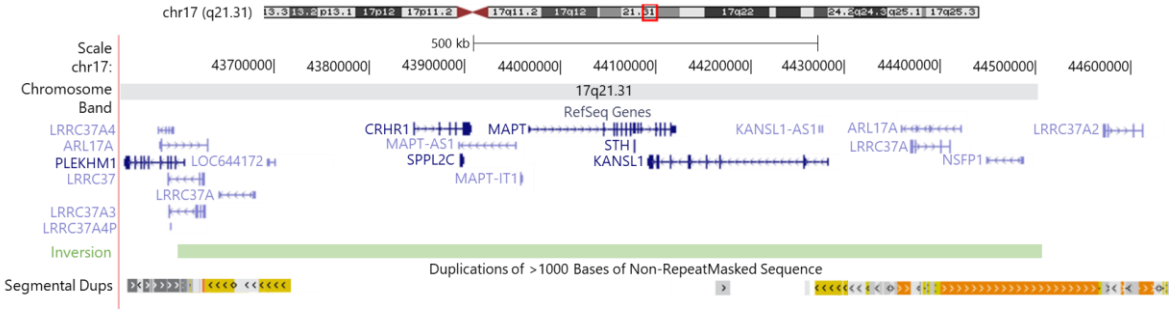


Figure 8. Schematic view of the 17q21.31 inversion. Represents the direct orientation corresponding to the H1 haplotype (hg19). Adapted from (150).

While H1 is commonly found in all human populations, H2 is mostly present in individuals with European ancestry with a frequency of ~20%. Also, evidence was found for positive selection limited only to European populations, concretely in Icelanders where carrier females were observed to have more children and higher recombination rates. In individuals from Africa and Asia, H2 haplotype was observed in 6-9% and 1%, respectively (139,148). Surprisingly,

comparative primate analysis suggested that H2 orientation represented the ancestral state of the inversion (151).

The 17q21.31 region contains multiple genes including *MAPT*, *CRHR1* (Corticotrophin Releasing Hormone Receptor 1) and *KANSL1* (KAT8 regulatory NSL complex subunit 1), among others. De Jong *et al.* studied the effect of the inversion status on gene expression within the region and found that several genes were functionally regulated by the inversion haplotypes. As other studies before, they found a higher expression of *MAPT* associated with the H1 status in the frontal cortex and the cerebellum. Expression of *PLEKHM1* (Pleckstrin Homology domain containing, family M Member 1) was also significantly increased in H1 in cerebellum. Conversely, H1 genotype was associated with a lower expression of *CRHR1* in cerebellum (152).

The increased *MAPT* expression associated with H1 haplotype leads to an overproduction of hyperphosphorylated protein Tau that aggregates in neurons. This mechanism has been linked to neurodegenerative disorders including progressive supranuclear palsy, corticobasal degeneration, Parkinson's disease and Alzheimer's disease (149,153–156). The H2 haplotype is related to recurrent microdeletion of the 17q21.31 region predisposing to the Koolen-De Vries syndrome, which is characterized by moderate ID, developmental delay, hypotonia and facial dysmorphism. It has been hypothesized that the haploinsufficiency of *MAPT* may affect axonal elongation and neuronal migration possibly explaining the major features of the syndrome (157,158).

3.2.3 Inversion 8p23.1

The 8p23.1 inversion is the largest inversion found in the human genome encompassing almost 4.7 Mb in chr8: 7,238,552-12,442,658 (hg19). It is flanked by two large blocks of SDs, the olfactory repeat regions REPD and REPP. Its estimated frequency varies depending on the population studied: ~52-59% in Africans, ~12-27% in Asians and ~26-50% in Europeans (139,144,159,160). The inverted region encompasses at least 50 genes and four of them showed statistically significant different expression levels. *NEIL2* (Nei Like DNA

Glycosylase 2) and *CTSB* (Cathepsin B) were reported to have a higher expression associated with the inverted allele, whereas *BLK* (B Lymphoid tyrosine Kinase) and *MSRA* (Methionine Sulfoxide Reductase A) genes were significantly decreased in the inverted conformation (160). Also, *PPP1R3B* (Protein Phosphatase 1 Regulatory subunit 3B) was found downregulated in inverted chromosomes in different tissues (144).

Genes encompassing this inversion, such as *PPP1R3B*, have been related to lipid metabolism and an association was found between the 8p23 inversion allele and low BMI (body mass index) in children of European ancestry (141,161). Also, variants in *FAM167A/BLK* locus, harboured in the inverted region, have been linked to risk of systemic lupus erythematosus (SLE) and other autoimmune diseases (162–165). Recent studies detected an association effect between the non-inverted status of inv8p23.1 and susceptibility to SLE (OR=1.18, $p=8.18x10^{-7}$) (166,167).

Carriers of the inverted 8p23.1 allele have an increased risk of having offspring with unbalanced chromosomal rearrangements including deletions and duplications as a result of NAHR events between REPD and REPP blocks (132). The 8p23.1 deletion syndrome is related to developmental delay, behavioural and cardiac problems, whereas the reciprocal duplication is associated with mild dysmorphism and developmental and speech delay (168). More complex interchromosomal rearrangements have been observed in this region, where olfactory receptor-gene clusters exist as well as in 4p16. This is why heterozygous individuals for both 8p23.1 and 4p16 inversion are at risk of originating an unbalanced t(4;8)(p16;p23) translocation in offspring because the heterozygous state allows the recombination of two non-homologous chromosomes. Subjects carrying der(4), which implies the deletion of 4p16 and the duplication of 8p23, show Wolf-Hirschhorn syndrome (OMIM #194190) characterized by growth deficiency, variable developmental delay, seizures and dysmorphism. On the other hand, individuals with der(8), which involves the duplication of 4p16 and the deletion of 8p23, have a less severe phenotype (169).

4. The 7q11.23 recurrent rearrangements

The human chromosome 7 exhibits an especially SD-rich sequence accounting for an 8.2% of the overall chromosomal content (170). Within this chromosome, there is one of the most well-studied complex genomic regions located in the 7q11.23 band, which has been thoroughly characterized. This region is composed of a unique copy sequence flanked by groups of LCRs that mediate NAHR events predisposing to deletions, duplications, triplications and inversions. CNVs in 7q11.23 region lead to neurodevelopmental disorders. The recurrent microdeletion causes WBS (OMIM #194050) characterized by cardiac and connective tissue problems, facial dysmorphism, ID and hypersociability. Conversely, the reciprocal duplication originates the 7q11.23 microduplication syndrome (OMIM #609757) which courses with speech delay and autism. All this points out that genes encompassed in the 7q11.23 region are dosage sensitive and may be involved in human speech and language (171).

4.1 7q11.23 genomic structure

The 7q11.23 region is composed of a ~1.2 Mb single-copy gene region flanked by three large groups of SDs known as centromeric (c), medial (m) and telomeric (t). At the same time, each of these LCRs comprises three differentiated blocks named A, B and C (Figure 9). Centromeric and medial SDs are identically orientated with different order, whereas the telomeric repeat is found in the same order as the centromeric one, but with inverted orientation (172,173). The majority of genes found in the different blocks are truncated copies or pseudogenes with the exception of some transcriptionally active and functional genes.

The high homology between flanking SDs, which is greater than ~97%, predisposes to misalignment and unequal crossing-over during meiosis leading to the 7q11.23 deletion or the reciprocal duplication. However, it is a highly dynamic region since polymorphic variation of LCR copy number and inversions have been observed in healthy population (135).

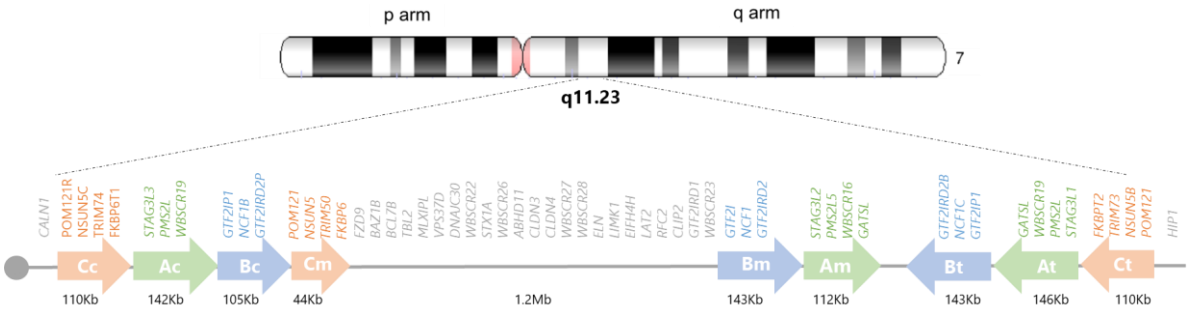


Figure 9. Schematic representation of 7q11.23 region. The ideogram of chromosome 7 is shown above. A detailed characterization of the region can be observed below. Centromeric (c), medial (m) and telomeric (t) SDs are represented as arrows which illustrate the order and orientation of the blocks. A block is shown in green, B block in blue and C block in orange. Adapted from (171).

4.2 Williams-Beuren syndrome

WBS is a rare neurodevelopmental and multisystem disorder caused by the hemizygous microdeletion of 26-28 contiguous coding genes encompassed in the 7q11.23 region with a prevalence of 1/7500 live births (174). Various individuals with unique facial features and infantile hypercalcemia were reported during the 1950s, but it was not until 1961 when Dr. Williams first recognized it to be a distinct clinical entity. Afterwards, Dr. Beuren reported more patients and since then it was known as Williams-Beuren syndrome (171,175,176).

4.2.1 Genetic mechanism

The 7q11.23 deletion arises when there is unequal crossing-over between directly oriented blocks in the centromeric and medial location (172,173). The size of the deletion varies depending on the LCRs implicated in the NAHR events, but two recurrent deletions are found among WBS individuals. The majority of cases (~90%) present a 1.55 Mb deletion due to recombination between the centromeric and medial B repeats (Bc/Bm) as they show 99.6% of sequence identity. In <10% of WBS individuals, the crossing-over is mediated by Ac and Am, which share 98.2% of sequence identity, resulting in a larger ~1.83 Mb deletion (Figure 10) (135). Atypical deletions have also been detected and account for 2-3% of the cases (177,178).

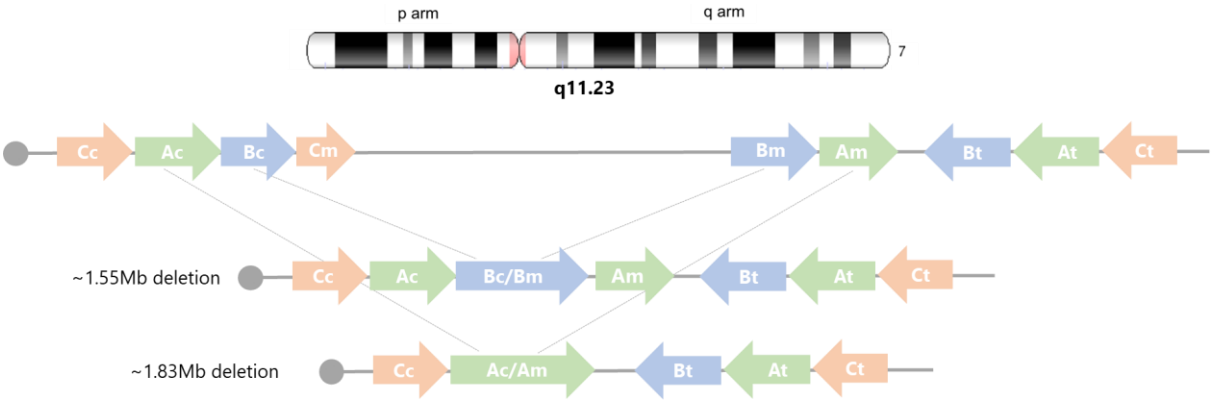


Figure 10. NAHR events generating the two recurrent deletions in WBS. Adapted from (171).

In two thirds of the WBS deletions, the NAHR event is interchromosomal (between different chromosomes 7s), whereas only one third is intrachromosomal (135). Although few cases of autosomal dominant inheritance have been reported, WBS usually occurs sporadically (179). Besides, no bias concerning the parental origin of the deletion has been observed (135).

The polymorphic 7q11.23 inversion has been identified in 25-28% of WBS transmitting progenitors, which compared to the rate in general population (5.8%) implicates an increased risk of having a child with WBS. Concretely, from a chance of 1/9500 for non-inversion carriers, the risk for parents heterozygous for the inversion rises to 1/1750 (135,180).

4.2.2 Clinical features of WBS

WBS patients show a wide range of clinical signs and symptoms including distinctive facial features, cardiovascular abnormalities, growth defects, gastrointestinal and endocrine anomalies, connective tissue problems, neurological problems and unique cognitive and behavioural profiles (181). The condition is usually diagnosed around the second year of life (182).

Regarding facial dysmorphism, WBS individuals present broad forehead, bitemporal narrowing, epicanthal folds, flat nasal bridge, short upturned nose, broad nasal tip, long philtrum, wide mouth with full lips, small jaw, widely spaced teeth, delicate chin and full cheeks (Figure 11) (171,181).

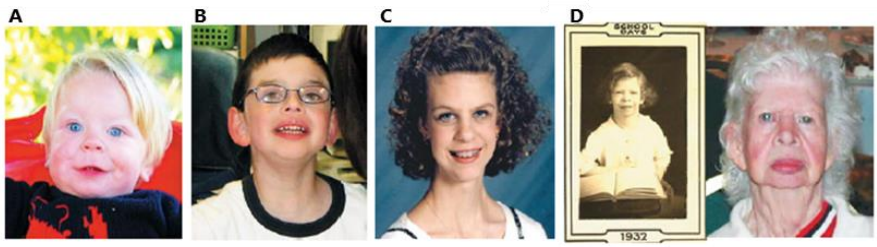


Figure 11. Distinctive facial appearance of subjects with WBS. Pictures of affected individuals at different ages: (A) young child, (B) school-age child, (C) young adult and (D) same subject at 12 years old in the left, and 83 years in the right. From (181).

The most frequent and clinically significant cardiovascular finding is supravalvular aortic stenosis (SVAS) found in approximately 64-70% of WBS patients. The haploinsufficiency of elastin gene (*ELN*) has been linked to the SVAS phenotype observed in WBS, but also with the connective tissue lesions. Other cardiovascular problems include pulmonary arterial stenosis (34%), hypertension (17-50%), mitral valve disease (15%) and ventricular sept defect (12%) (181–183). Cardiovascular complications are considered the major cause of death among patients with WBS. In fact, the risk of sudden cardiac death is 25 to 100 times higher in WBS compared to general population (184).

During infancy, typically observed symptoms are feeding problems and failure to thrive in 70% and 80% of WBS individuals, respectively. Abnormal weight gain and short stature have also been reported. In relation to gastrointestinal anomalies, constipation, vomiting, colic, diverticular disease, chronic abdominal pain and gastroesophageal reflux are common features in WBS (182,185). Some of these symptoms are directly related to hypercalcemia, which can be found both during infancy and adulthood, but affecting a minority of patients (6-15%) (171). Other endocrine anomalies are hypercalciuria in most patients, high prevalence of diabetes mellitus and subclinical hypothyroidism in 15-30% of cases (181,185). As for connective tissue alterations, hoarse voice, hernias, joint laxity and soft skin are frequently seen in WBS subjects (171).

Coordination difficulties, infantile hypotonia, hyperreflexia, cerebellar dysfunctions, strabismus, nystagmus and sensorineural hearing loss are some of the neurological problems observed in individuals with WBS. Also, around

85% of individuals develop hypersensitivity to certain noises and only ~10% of cases suffer from seizures (171,182).

Concerning the cognitive profile, all WBS patients present developmental delay and mild to severe ID with an average IQ (Intelligence Quotient) of 55-62 in about 75%. Also, around 25% show language delays and they have a consistent pattern of cognitive strengths and weaknesses. Auditory rote memory and language ability are better than expected, whereas visuospatial construction and visual perception are severely impaired. They tend to focus on the details and do not appreciate the whole picture (181,186). As for the behavioural phenotype, WBS patients show a highly sociable, approaching and overfriendly personality with good social and interpersonal skills. They are extremely empathic persons and driven to engage with strangers, which may put them on risk for harm (186,187). However, individuals with WBS manifest high levels of anxiety, excessive fears and phobias as well as generalized worries and irritability. Moreover, ADHD was reported in 65-84% of cases (182,188).

Although WBS and ASD have repeatedly been described as opposite disorders, several cases of WBS with comorbid ASD features have been reported to date in the literature (189–195). Recently, a meta-analysis studying the comorbidity of ASD with a variety of genetic and metabolic syndromes found that 12% of WBS individuals have coexisting ASD (196). This represents a 6-10-fold increase in the ASD prevalence in WBS compared to general population. In fact, some socio-communicative impairments have been identified in WBS patients including delayed acquisition of language, difficulties with language pragmatics and non-verbal aspects of communication, delays in pointing or showing objects, problems differentiating lies and jokes, repetitive behaviours and difficulties establishing and maintaining friendships. Therefore, this pattern of socio-communicative deficits overlaps with those seen in children with ASD (197).

4.3 7q11.23 microduplication syndrome

The 7q11.23 microduplication syndrome is caused by the reciprocal heterozygous duplication of the WBS critical region. As the WBS deletion was originated by NAHR between flanking LCRs, it was hypothesized that duplications should also occur, but it was not until 2005 when the first duplication case was reported (198). Deletions and duplications are equal in size, commonly 1.55-1.83 Mb, encompassing 28-30 genes (199). The estimated frequency of the duplication was 1/7500–20000 individuals with higher parental genetic transmission compared to WBS deletion (200,201).

Usually, duplications have milder consequences than deletions, which explains why individuals with the 7q11.23 microduplication, or Dup7, have a milder phenotype that is less distinct and more variable than in WBS (171). In addition, this is also supported by the fact that a triplication case of the WBS region was reported in a patient with severe speech delay (202). Patients with the 7q11.23 microduplication syndrome show mildly dysmorphic features, cardiovascular defects, growth and development anomalies, neurological problems and specific cognitive and behavioural profiles. For a comparison with WBS features see table 4.

The characteristic facial features of Dup7 individuals include broad forehead, asymmetric face, macrocephaly, abnormal head shape, straight eyebrows, deep-set eyes, high broad nose, short philtrum, thin upper lip, high arched palate and over folded helices (Figure 12) (203,204).

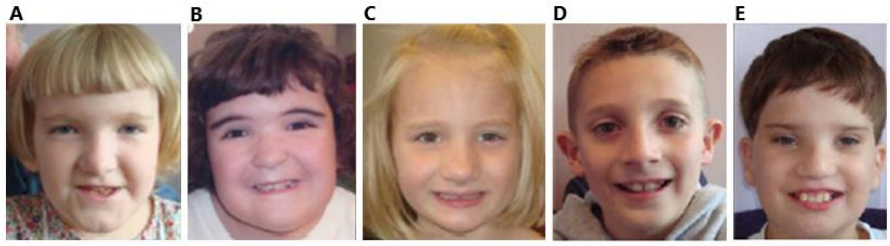


Figure 12. Facial features of Dup7 individuals. Pictures of affected individuals at mild childhood: (A) 5.7 years, (B) 5.9 years, (C) 6.8 years, (D) 8 years, (E) 10.3 years. Adapted from (203).

Regarding cardiovascular anomalies, the most common finding is aortic dilation observed in 30-46% of patients. Other defects comprise patent ductus arteriosus, atrial and ventricular septal defects and subvalvular aortic stenosis (203,204). Growth and development abnormalities include developmental delay (70%), macrocephaly (50%), short stature (13-17%) and obesity (18%). Among neurological problems, in 48-58% of cases hypotonia was reported, whereas seizures in ~20%. Abnormal brain imaging, such as decreased cerebral white matter volume, mild cerebral atrophy or ventriculomegaly, was found in 80% of Dup7 individuals and deficits of gait and station suggesting cerebellar dysfunction were observed in 62% of cases (199,203).

In relation to the cognitive profile, 91% of Dup7 patients show significant speech and language delay, which is considered a hallmark feature of this syndrome. Usually, expressive language tends to be more impaired than receptive language, and symptoms can be divided into oral or verbal apraxia, phonological defects and dysarthria (204). In fact, Speech Sound disorder was diagnosed in 82.5% of Dup7 subjects (203,205). In order to compensate this deficit, the majority of individuals develop non-verbal communication including signs, gestures, pointing, drawing or writing (171). It has been reported from severe ID to high average nonverbal intellectual ability among Dup7 patients, but around 38% present mild-to-severe ID (204,205).

Concerning the behavioural profile, anxiety disorders were the most common psychiatric finding among Dup7 individuals including social anxiety (42-50%), selective mutism (29%), separation anxiety disorder (13%) and specific phobia (36-53%). ADHD and oppositional defiant disorder were diagnosed in 37% and 24% of cases, respectively (204,205). Other behavioural phenotypes include aggressiveness, temper outbursts, stereotypical movements, sensory issues, sleep problems and self-injury (199).

The 7q11.23 microduplication has been significantly associated with ASD (84). Autistic features such as poor social interaction, poor eye contact, limited facial expressions and repetitive behaviours have been described in Dup7 individuals. A recent study focusing on ASD symptomatology among patients carrying the

7q11.23 duplication found a consistent ASD diagnosis in 19% of cases and 39.7% when taking into account parental reports. Therefore, there is a higher prevalence of ASD among Dup7 individuals and the 7q11.23 duplication is considered a risk factor for ASD (194). Also, this duplication confers a 10-fold increase in risk for schizophrenia (206).

Lastly, one or more congenital anomalies have been reported in approximately 30% of individuals with the 7q11.23 duplication. Other medical problems reported are cryptorchidism (10%), cutis marmorata (24%), joint laxity (19%), growth hormone deficiency (9%), *café au lait* spots (2%), enuresis (2%), hydrocephalus, strabismus and lordosis (171,203,204).

Feature	7q11.23 microdeletion	7q11.23 microduplication
Craniofacial	Broad forehead Periorbital fullness Low nasal root Full lips Long philtrum	Broad forehead Deep set eyes High, broad nose Thin upper lip Short philtrum
Growth and development	Delayed growth, failure to thrive Developmental delay	Normal growth Developmental delay
Endocrine	Hypercalcemia	Normal calcium levels
Musculoskeletal	Joint laxity	Joint laxity
Cardiovascular	Supravalvular aortic stenosis	Aortic dilation
Neurological and cognitive	Relative strength in expressive language Hypotonia Poor visuospatial skills Mild intellectual disability	Speech and language delay Hypotonia Poor visuospatial skills Mild intellectual disability
Psychiatric/behaviour	ADHD Lack of stranger anxiety Autism Specific phobias No aggression	ADHD Social anxiety Autism Specific phobias Aggression/temper outbursts

Table 4. Comparison of clinical features observed in the 7q11.23 deletion causing WBS and the reciprocal microduplication. ADHD: Attention Deficit Hyperactivity Disorder. Adapted from (171,204).

5. Molecular pathophysiology of ASD

The molecular bases of ASD continue to be largely unknown. Many large-scale genetic studies, including GWAS, arrays and NGS strategies, have been useful in the identification of ASD risk genes. In order to uncover the deregulated pathways underlying ASD, functional and pathways enrichment tests are commonly used approaches (49,207). Concretely, ASD risk genes encode proteins that can be grouped in two major biological classes: synapse processes and transcription/chromatin modification (Table 5) (105,208).

Function	Genes associated with ASD
Synaptic gene transcription and translation pathway	*FMR1, TSC1, TSC2, PTEN, NF1, CYFIP1, EIF4E, EIF4EBP2, SYNGAP1*
Synaptic cell adhesion molecules	Neurexin and Neuroligin families, *CNTN4, CNTN6, PCDH10*
Neurotransmitters and scaffolding proteins in synapse	*SHANK1, SHANK2, SHANK3, DLG4, GPHN*
Ion channel and transport proteins	*CACNA1A, CACNA1E, CACNA1H, CACNB2, SCN1A, SCN2A, KCNQ3, KCNQ5, KCND2*
Chromatin remodelling and transcription regulation	*CHD8, BAF155, MECP2, MEF2C, ADNP, POGZ*
Ubiquitination pathway	*UBE3A, PARK2, TRIM33*

Table 5. Functions of ASD-associated genes. Adapted from(49,208,209).

There are multiple synaptic aspects affected by ASD risk genes including synapse formation and elimination, synaptic transmission and plasticity. This suggests that synaptic dysfunction is partially involved in the ASD pathogenesis (209). These ASD risk genes have highlighted several disrupted signaling pathways including glutamatergic synapse, WNT/β-catenin signaling, PI3K/AKT/mTOR pathway, RAS/ERK pathway and MAPK signaling pathway (88,210,211). The majority of the pathways mentioned are highly integrated as can be seen in figure 13.

Taking all together, ASD risk genes seem to be implicated in many aspects of basic cell function ranging from basic metabolism, transcription regulation, synaptic transmission and mRNA translation to neurogenesis, neuronal

migration, axon guidance or dendrite outgrowth (207). In addition, there is an increasing number of genes involved in chromatin structure and epigenetic regulation (210).

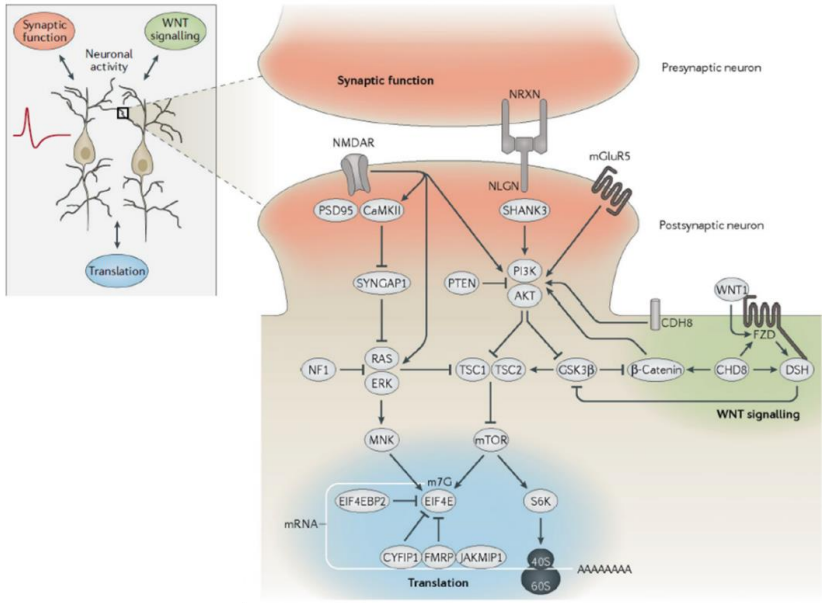


Figure 13. Major cellular pathways affected in ASD. Adapted from (212).

Transcriptomic studies, including microarray gene expression and RNA sequencing (RNAseq), have been crucial approaches for the investigation of molecular pathways and networks related to ASD. Although the majority of transcriptomic studies have been performed in blood tissues, there are some that have been carried out in post-mortem brain tissues (213). The earliest microarray study in brain focused on the gene expression pattern of targeted genes relevant to ASD finding an upregulation of genes involved in glutamate signaling (214). This alteration could cause an excitatory-to-inhibitory imbalance that has been hypothesized to underlie ASD (215).

Later on, weighted gene co-expression network analysis (WCGNA) were implemented (213). Voineagu *et al.* performed a WGCNA analysis in transcriptomic data of 19 ASD patients and 17 controls identifying two modules. The downregulated module was enriched for synaptic function, vesicular transport and neuronal projection, whereas the upregulated module was

enriched for astrocyte and activated microglial markers as well as immune system and inflammatory response (216). This study supported previous findings from smaller studies (217) and posteriorly has been replicated in larger independent samples (218,219). Collectively, these studies found transcripts involved in neuronal functions and synaptic signaling to be under-expressed in ASD brains, while transcripts related to immunologic and microglial functions were up-regulated pointing out the interplay between the innate immune system and neuronal activity in ASD (213). RNAseq studies in ASD patient-derived neurons showed similar results emphasizing the involvement of axon guidance, glutamate receptor signaling and calcium signaling, among others (220). Curiously, in a recent mega-analysis of blood transcriptomes using WGCNA, the characteristic ASD blood signature identified was reduced expression of transcripts related to innate immune and inflammatory signaling. This discordance could be explained by differences in the tissue where the inflammation originates, by post-mortem hypoxic damage or by tissue-related differences in negative feedback response (221).

Another recurrent finding among transcriptomic studies is the aberrant number of differentially expressed lncRNAs both in brain and blood tissues of ASD patients, which are transcripts greater than 200 nucleotides in length that do not encode for proteins (219,222,223). LncRNAs are considered key components of various cellular processes with control of every level of the regulation in gene expression pathways including transcriptional, post-transcriptional and epigenetic regulation (224). In addition, microRNAs (miRNAs) have also been described to be deregulated in ASD individuals and to be implicated in neurodevelopmental processes disrupted in ASD (212).

Other strategies to assess deregulated pathways in ASD are the generation of physical interaction data, from experimental design or curated literature such as protein-protein interactions (PPI) network (213). One study applying PPI network analysis in *de novo* variants from exome sequencing data found that genes with disruptive mutations mapped to a highly interconnected β-catenin/chromatin remodeling protein network (225). Subsequent studies using

PPI network analysis considered both *de novo* and inherited mutations and found four clusters: cell junction and TFG-β pathway, cell communication and synaptic transmission, neurodegeneration and transcriptional regulation (105). Another strategy to construct networks is the Network-Based Analysis of Genes (NETBAG) approach, which combines multiple forms of literature-curated data to identify network modules reflecting phenotypes (213). The first study using NETBAG used genes encompassed in CNVs related to ASD and found a highly interconnected module linked to synaptic function as well as chromatin remodeling, transcription regulation and MAPK signaling (Figure 14) (210).

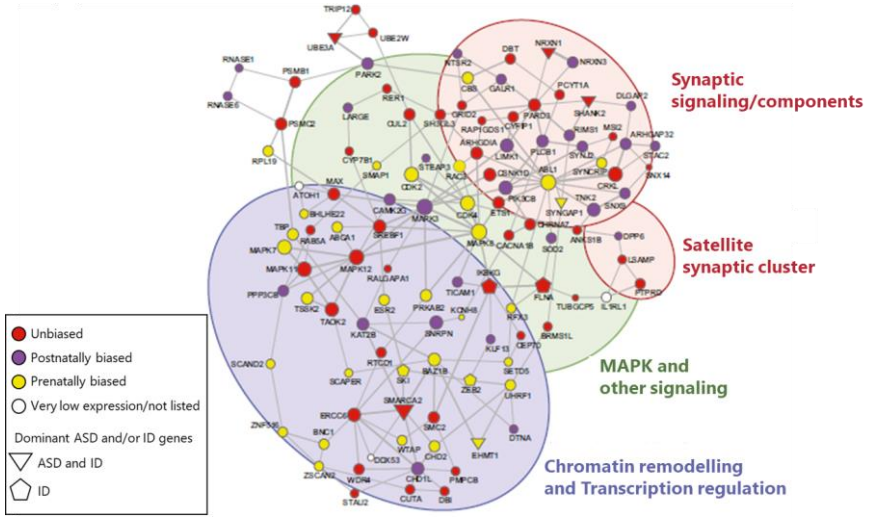


Figure 14. NETBAG results from the analysis of genes encompassed in rare *de novo* CNVs. Nodes correspond to genes and edges to interactions. Colours indicate whether genes show prenatal or postnatal-biased gene expression: red for unbiased, purple for postnatally biased, yellow for prenatally biased and white for unlisted genes. Adapted from (210).

ASD has been hypothesized to be a multistage, progressive disorder of brain development. There is evidence to support that it spans from prenatal to early post-natal life. In fact, deregulated pathways are crucial for fetal and early post-natal stages starting with dysregulation of cell proliferation and differentiation and ending with disruption of synaptogenesis and reduced neural network functioning (211).

OBJECTIVES

Objectives

The main goal of this thesis is to explore complex genomic regions and its implications for ASD as well as to deepen in the identification of deregulated pathophysiological mechanisms in ASD. This work could not have been possible without the generosity of around 300 families affected with ASD in giving samples and accessing clinical information as well as available genetic data in publicly accessible servers.

In order to achieve this goal, we established the following specific objectives:

1. To investigate the known 7q11.23 complex region through the assessment of genetic factors contributing to ASD in WBS patients.

2. To study the transcriptomic consequences of 7q11.23 complex region using iPSC-based models from Dup7 and WBS individuals.

3. To explore the role of two unexplored complex regions of the genome, ancestral inversions 17q21.31 and 8p23.1, and to evaluate its functional consequences through transcriptomic analysis.

4. To identify the presence of genes that recurrently harbour rare genetic variants in ASD individuals and to elucidate common deregulated functional pathways.

CHAPTER 1

Marta Codina-Sola*, **Mar Costa-Roger***, Debora Pérez-García, Raquel Flores, Maria Gabriela Palacios-Verdú, Ivon Cusco and Luis A Pérez-Jurado. **Genetic factors contributing to autism spectrum disorder in Williams-Beuren syndrome**. *J Med Genet* 2019; 0: 1-8.

https://jmg.bmj.com/content/56/12/801.long

PMID: 31413120 DOI: 10.1136/jmedgenet-2019-106080

*Joint first authors.

Contributions:

MC-S and MC-R analyzed the exome data, performed the experiments and drafted the manuscript. DP-G and MGP-V contributed with the clinical and phenotypic information. RF contributed with detailed molecular characterization of deletions. IC and LAP-J conceived the study and participated in the design and data interpretation, and helped in drafting the manuscript. All authors read and approved the manuscript.

Cognitive and behavioural genetics

ORIGINAL ARTICLE

Genetic factors contributing to autism spectrum disorder in Williams-Beuren syndrome

Marta Codina-Sola,[1,2] Mar Costa-Roger,[1] Debora Pérez-García,[1] Raquel Flores,[1] Maria Gabriela Palacios-Verdú,[1,3] Ivon Cusco,[1,2] Luis Alberto Pérez-Jurado[1,4]

► Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmedgenet-2019-106080).

For numbered affiliations see end of article.

Correspondence to
Dr Ivon Cusco, Clinical and Molecular Genetics Area, Vall Hebrón Hospital Research Institute (VHIR), Hospital de la Vall d'Hebron, Hospital Vall d'Hebron, Barcelona 08035, Spain; icusco@vhebron.net

MC-S and MC-R are joint first authors.

Received 13 February 2019
Revised 11 July 2019
Accepted 16 July 2019

To cite: Codina-Sola M, Costa-Roger M, Pérez-García D, *et al*. *J Med Genet* Epub ahead of print: [please include Day Month Year]. doi:10.1136/jmedgenet-2019-106080

ABSTRACT

Background The hallmark of the neurobehavioural phenotype of Williams-Beuren syndrome (WBS) is increased sociability and relatively preserved language skills, often described as opposite to autism spectrum disorders (ASD). However, the prevalence of ASD in WBS is 6–10 times higher than in the general population. We have investigated the genetic factors that could contribute to the ASD phenotype in individuals with WBS.

Methods We studied four males and four females with WBS and a confirmed diagnosis of ASD by the Autism Diagnostic Interview-Revised. We performed a detailed molecular characterisation of the deletion and searched for genomic variants using exome sequencing.

Results A de novo deletion of 1.55 Mb (6 cases) or 1.83 Mb (2 cases) at 7q11.23 was detected, being in 7/8 patients of paternal origin. No common breakpoint, deletion mechanism or size was found. Two cases were hemizygous for the rare T allele at rs12539160 in *MLXIPL*, previously associated with ASD. Inherited rare variants in ASD-related or functionally constrained genes and a de novo nonsense mutation in the *UBR5* gene were identified in six cases, with higher burden in females compared with males (p=0.016).

Conclusions The increased susceptibility to ASD in patients with WBS might be due to additive effects of the common WBS deletion, inherited and de novo rare sequence variants in ASD-related genes elsewhere in the genome, with higher burden of deleterious mutations required for females, and possible hypomorphic variants in the hemizygous allele or *cis*-acting mechanisms on imprinting.

INTRODUCTION

Williams-Beuren syndrome (WBS, OMIM#194050) is a rare neurodevelopmental disorder resulting from an heterozygous deletion of 25–27 genes at chromosome 7q11.23, estimated to affect approximately 1 in 7500 individuals.[1] The WBS locus has a complex genomic architecture with a single copy region flanked by three large segmental duplications, each composed of three major blocks (A, B and C), located in the centromeric (c), medial (m) and telomeric (t) sides.[2] Two types of recurrent rearrangements promoted by non-allelic homologous recombination (NAHR) can result in WBS syndrome. The most frequent is a 1.55 Mb deletion occurring between the medial and centromeric B blocks in 87% of cases. Around 10% of patients present a larger 1.83 Mb deletion due to a crossing over between centromeric and medial A blocks. The remaining 3% of patients with WBS show atypical deletions mediated by other mechanisms.[3 4] The WBS multisystemic phenotype is characterised by cardiovascular disease, distinctive facies, connective tissue abnormalities and growth and endocrine alterations, among others.[2 5] The main neurobehavioural hallmarks are mild to moderate intellectual disability (ID), hypersociability and relative language preservation.[6] Interestingly, the reciprocal duplication of 7q11.23 (OMIM#609757) results in a phenotype with speech delay, language impairment, milder learning problems and clear social interaction deficits often associated with autism spectrum disorder (ASD).[6 7]

Due to their associated behavioural manifestations, ASD and WBS have often been described as diametric opposite disorders, although this consideration is an oversimplification of both phenotypes.[8 9] Moreover, several cases of WBS with comorbid ASD have been reported and some authors have suggested that ASD features should be considered as part of the WBS phenotype.[7–18] A recent meta-analysis of the comorbidity of ASD features in several well-defined genetic syndromes concluded that the prevalence of ASD features among WBS individuals is as high as 12%.[19] Therefore, the frequency of ASD is 6–10 times higher in individuals with WBS than in the general population, a striking finding considering the typical WBS neurocognitive profile.

Potential modifiers of the common neurobehavioural phenotype of WBS include: (1) *cis*-acting mechanisms due to variable breakpoints altering flanking genes, although most patients described present common deletions of identical size[3 17]; (2) *trans*-acting factors present in the non-deleted hemizygous allele; (3) genetic mutations and/or structural variants elsewhere in the genome; (4) environmental events with or without epigenetic effects, including medical complications during development and early life. All these factors can have additive effects acting on a sensitised background caused by haploinsufficiency at the WBS locus.[10 17] Two individuals with WBS and co-occurring ASD previously reported presented hyperserotonaemia and were homozygous for the *short* (s) allele in the promoter of the serotonin transporter *SLC6A4* (5-HTTLPR), suggesting a possible modifier role of this locus.[14 17]

In the present work, we have investigated the genetic factors that could contribute to the ASD phenotype of eight individuals with WBS and

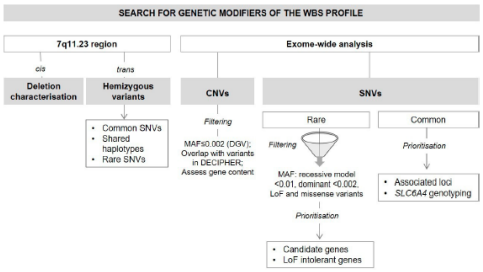


Figure 1 Strategy followed for the identification of second-hit genetic factors. DGV, Database of Genomic Variants; LoF, loss of function; MAF, minor allele frequency; SNV, single nucleotide variant.

comorbid ASD. We completed a detailed molecular characterisation of the deletion, genotyped the reported polymorphism at *SLC6A4* and performed a genome-wide unbiased search of second-hits by exome sequencing following the strategy depicted in figure 1.

METHODS

Patient selection

From a cohort of 122 individuals with a diagnosis of WBS and confirmed 7q11.23 deletion by molecular techniques, we selected four males and four females aged 6–31 years with an associated diagnosis of ASD. The diagnosis was based on the direct observation by a trained psychologist and clinicians and confirmed in all of them using the Autism Diagnostic Interview-Revised (ADI-R). Written informed consent was obtained from all parents or legal caregivers.

7q11.23 deletion characterisation

Blood samples from probands and parents were obtained and genomic DNA was extracted using the Puregene DNA Purification Kit (Gentra Systems, Big Lake, Minnesota, USA). The size and parental origin of the deletion was established by the analysis of multiple ligation-dependent probe amplification and several single and multiple-copy microsatellites. Refined mapping of deletion breakpoints was also performed by quantitative analysis of paralogous sequence variants as previously described.[3]

SLC6A4 genotyping

The polymorphism in the serotonin transporter promoter *SLC6A4* (5-HTTLPR) was genotyped by PCR and agarose gel electrophoresis, using primers previously described.[17]

Exome sequencing and analysis

Exomes were captured using the SureSelect Human All Exon V5 capture kit (Agilent, Santa Clara, California, USA) and libraries were sequenced on an Illumina MiSeq platform. Paired-end sequences were obtained with a read length of 250 bp.

Mapping, variant calling and filtering were performed using BWA and GATK's standard parameters. The hg19 human genome reference version was used. Variant annotation was performed using ANNOVAR (http://www.openbioinformatics.org/annovar/), considering 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.

For CNV detection, we applied ExomeDepth and compared our samples with a matched aggregate reference set of 248 in-house exomes captured and sequenced using the same protocol. CNVs were filtered based on their overlap with variants previously described in the Database of Genomic Variants and DECIPHER.[20]

Rare variant analysis and validation

We selected exonic variants with a minor allele frequency (MAF) lower than 0.002 according to several databases (previously mentioned) for heterozygous variants following a dominant inheritance model. For homozygous or compound heterozygous variants, under a recessive inheritance model, we selected a MAF ≤0.01. Since second-hit variants would act in the presence of a major hit and would not be expected to solely cause ASD, 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 for homozygous or compound heterozygous variants at ≤0.01.

To validate variants and perform segregation studies in parental samples, we used Sanger sequencing by capillary electrophoresis (ABI PRISM 7900HT, Applied Biosystems, Foster City, California, USA). Primers were designed with the PRIMER3 program (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/) and PCR reactions were carried following standard conditions.

Epistatic effects of rare variants

To study the putative contribution of epistatic effects of rare SNVs on gene expression deregulation in WBS, we compiled a list of genes previously altered in WBS by various mechanisms, including transcriptional dysregulation, differential methylation and the direct *GTF2I* targets. We performed a systematic literature review and selected only high-quality studies done in human subjects, obtaining 2251 candidate genes (online supplementary table 1).[21–24]

Analysis of ASD susceptibility loci

To study the potential contribution of common variants, we selected loci previously associated with ASD by Genome-Wide Association Studies (GWAS). Variants containing the term 'Asperger' or 'Autism' were extracted from the Genome-Wide Repository of Associations between SNPs and phenotypes (GRASP) database V.2.0 (online supplementary table 2).[25] Allele frequency in the WBS cohort was then compared with non-Finnish European (NFE) population data from ExAC by Fisher's exact test and q-value and false discovery rate (FDR) values were calculated using the R package qvalue.[26] To avoid population stratification, ExAC allele frequencies were compared with that reported in Spanish Variant Server, including data from 578 Spanish individuals (http://csvs.babelomics.org/).

RESULTS

Clinical characteristic of selected patients

The study was performed in eight Spanish patients with WBS (four males and four females), which had periodic follow-up in

Table 1 Clinical characteristics of the WBS individuals with associated ASD

	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
Relevant family history	No	No	Yes*	No	No	Yes†	No	No
Age at WBS diagnosis	7 y	3 y	5 m	3 m	3 m	2 y	5 y	1 y
Age at ASD diagnosis	4 y	10 y	6 y	5 y	12 y	5 y	3 y	5 y
ADI-R								
A	22	11	11	22	14	10	19	18
B (V)	–	7	4	–	–	–	–	–
B (NV)	10	–	–	12	7	4	14	9
C	8	4	6	8	8	6	9	5
D	5	4	1	5	3	4	4	5
IQ (WISC-R)			40		<40	41	<40	
Other neurological symptoms		ADHD	ADHD		Behavioural problems	Epilepsy, ADHD		ADHD
Cardiovascular disease	ND	SVAS	Coarctation of aorta	SVAS	SVAS	SVAS	SVAS	ND
Endocrine abnormalities	ND	ND	ND	ND	Early puberty	ND	Early puberty	Subclinical hypothyroidism

ADI-R scores: A. Social interaction (cut-off: 10); B. Communication and language, (V): verbal (cut-off: 8), (NV): non-verbal (cut-off: 7); C. Restricted and repetitive behaviours (cut-off: 3); D. Developmental alterations earlier than 36 months (cut-off: 1).
*Paternal aunt with ASD and severe ID.
†Father committed suicide.
ADHD, attention deficit hyperactivity disorder; ADI-R, Autism Diagnostic Interview-Revised; ASD, autism spectrum disorder; F, female; IQ, intellectual quotient; M, male; m, month; ND, not detected by standard testing; SVAS, supravalvular aortic stenosis; WBS, Williams-Beuren syndrome; WISC-R, Wechsler Intelligence Scale for Children-Revised; y, year.

a multidisciplinary clinic and complete neurobehavioural evaluation due to a comorbid diagnosis of ASD. Their main clinical characteristics are summarised in table 1. All patients fulfilled criteria in the social interaction and restrictive behaviour domains while two of them (WBS2 and WBS3) did not reach the verbal communication domain threshold. Family history of ASD or psychiatric disorders was positive on the paternal side in two cases (WBS3 and WBS6) and negative in the others (table 1).

Detailed characterisation of the 7q11.23 deletion

To define potential *cis*-acting factors in the deleted allele, we characterised all deletions at the molecular level including determination of the parental origin. All rearrangements were de novo and all but one (7/8) were of paternal origin (table 2). The most common 1.55 Mb deletion was found in 6/8 patients, whereas 2/8 carried the larger recurrent 1.83 Mb deletion mediated by NAHR between A blocks (figure 2). In two patients (WBS1 and WBS3), 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 WBS cognitive phenotype,[27] we analysed the deletion breakpoints to assess the number of functional copies of this gene in the patients (figure 2). As expected, the deletions mediated by A blocks included the entire medial B block with a copy of *GTF2IRD2*, resulting in a loss of one functional copy (1M/2T: 1 medial (*GTF2IRD2*)/2 telomeric (*GTF2IRD2B*)). The two

Table 2 Summary of the main genetic findings per patient

	7q11.23 deletion			Hemizygous allele	Genome-wide	Genome-wide rare LoF SNVs	
Individual	Origin	*GTF2IRD2*	Size (Mb)	SNVs	CNVs	ASD candidate genes	LoF intolerant genes
WBS1	Pat	1M/3T	1.55				
WBS2	Pat	2M/2T	1.55	T at rs12539160			
WBS3	Pat	1M/3T	1.55		*SIK1 partial dup*^{PI}		
WBS4	Pat	1M/2T	1.83			*CC2D1A* c.1357-2A>C^{PI}	
WBS5	Pat	2M/2T	1.55			*PYHIN1* p.(Arg373*)^{MI} *UBR5* p.(Arg633*)^{DN}	*SEC24C* c.2682+2A>G^{MI}
WBS6	Pat	1Ch/1M/2T	1.55	T at rs12539160		*AGAP1* c.1051-2A>T^{PI}	
WBS7	Mat	2M/2T	1.55		*DUSP22* del^{NM}	*USP45* p.(Glu220*)^{MI}	*CXXC1* p.(Ser222Leufs*7)^{NM}
WBS8	Pat	1M/2T	1.83			*PIK3CG* p.(Glu14Glyfs*147)^{PI}	*MED26* c.147+2T>C^{PI} EPHB1 p.(Ser435ProfsTer13)^{MI}

*Stop codon.
Ch, chimeric-type copies; DN, de novo; del, deletion; dup, duplication; LoF, loss of function; M, medial-type copies; MI, maternally inherited; Mat, de novo, maternal chromosome; NM, not maternal (father not available); PI, paternally inherited; Pat, de novo, paternal chromosome; T, telomeric-type copies; WBS, Williams-Beuren syndrome.

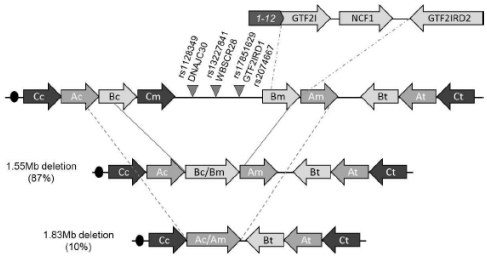


Figure 2 Schematic representation of the Williams-Beuren syndrome (WBS) locus, showing the two most common deletions and the gene content of the B block. The B block contains three genes: *GTF2I*, *NCF1* and *GTF2IRD2*. Whereas *GTF2I* and *NCF1* have a single functional copy located at the medial B block, *GTF2IRD2* has two functional copies, located at the medial (*GTF2IRD2*) and telomeric (*GTF2IRD2B*) B blocks. The 1.55 Mb deletion is mediated by B blocks and results in a chimeric medial-centromeric block, with the number of functional copies of *GTF2IRD2* and *NCF1* depending on the deletion breakpoint. In contrast, the 1.83 Mb deletion is mediated by A blocks, resulting in the loss of the medial and centromeric B blocks with functional copies of *GTF2IRD2* and *NCF1*.

patients with inversion-mediated deletions had a 1M/3T genotype, originated from the loss of a medial copy and a gain of a telomeric copy of *GTF2IRD2*. Among the patients with 1.55 Mb deletions, three had breakpoints before *GTF2IRD2*, with no change in the number of functional copies (2M/2T), whereas in one patient the breakpoint occurred within *GTF2IRD2*, creating a chimeric copy (Ch) between the medial gene and the centromeric pseudogene (1Ch/1M/2T) (table 2). The breakpoints also affect the number of *NCF1* copies, by either deleting or not one of the functional copies.

Hemizygous variants in the 7q11.23 region

To look for *trans*-acting factors in the 7q11.23 allele present in hemizygosity, we analysed the entire captured region within the WBS common deletion locus (chr7:72 700 000–74 250 000) looking for over-represented rare and common variants, shared haplotypes and rare deleterious SNVs. For over-represented variants, we compared allele frequencies of all described hemizygous variants (n=32) in our cohort to those reported in ExAC for European population. Variants with significantly different frequencies in ExAC and Spanish Variant Server were excluded to avoid population stratification. Due to the small sample, none of the variants reached statistical significance, but we identified a nearly significant association (p=0.076) with rs12539160, which was present in two individuals (WBS2 and WBS6). This synonymous variant located near an exon-intron boundary of *MLXIPL* had been previously associated with ASD in a GWAS.[28] Taking advantage of the hemizygosity of SNPs in the deleted single-copy region, we extracted phased haplotypes in this 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*). 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 the WBS locus. Only hemizygous 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 non-synonymous variants in genes *MLXIPL* and *TBL2* predicted as tolerant by various protein effect prediction algorithms ((SIFT (sorting intolerant from Tolerant), PolyPhen (Polymorphism Phenotyping) and Condel)). No rare variants were identified at *GTF2I*, *NCF1* and *GTF2IRD2*, the genes in the flanking segmental duplications.

Genome-wide analysis of rare variants

Copy number variants

We also studied the presence of additional rearrangements that could explain the autistic symptoms in our cohort. In addition to the WBS 7q11.23 deletion, we observed an average of 25 CNVs per patient ranging from 170 bp to 334 kb. None of the additional rearrangements overlapped with known genomic disorders or was previously associated with neurodevelopmental disorders and all overlapped with previous CNVs described in the general population.[29] Only two CNVs comprised ASD candidate genes (*SIK1* and *DUSP22*) (online supplementary table 3).[30] The CNV involving *SIK1* was a partial duplication affecting the last two exons of the gene, paternally inherited in patient WBS3. The heterozygous deletion completely containing *DUSP22* was found in patient WBS7 and not in her mother, but paternal sample was unavailable.

Rare single nucleotide variants

We prioritised deleterious variants present in a list of candidate genes from -Simons Foundation Autism Research Initiative (SFARI) (n=791) (online supplementary table 3) and/or highly constrained genes, as previous studies have shown that genes predisposing to ASD carry a low burden of disrupting mutations in the general population.[30 31] Constrained genes were defined as those with a probability of being loss of function (LoF) intolerant (pLI >0.9) according to ExAC.[32] LoF and missense variants predicted as damaging by both SIFT and PolyPhen were considered deleterious variants.

We detected a total of 38 rare deleterious SNVs in 34 candidate genes (online supplementary table 4). All LoF variants (n=6) were validated and segregation studies were performed on all samples, showing a de novo variant in the *UBR5* gene in case WBS5 (table 2). De novo variants in the *UBR5* gene, two missense and one LoF, have been previously described in ASD individuals.[33–35] WBS5 is a severely affected female who presented an obsessive ritualistic behaviour, hypersensitivity to noise and physical stimuli, had not developed language and did not show eye contact.

As for highly constrained genes, excluding candidate genes, we detected 68 SNVs in 65 genes, including four LoF mutations (table 2). All genes selected on the basis of being LoF intolerant but harbouring a LoF in our cohort were brain-expressed and some had been previously associated with psychiatric and neurodevelopmental disorders, such as *SEC24C*, *CXXC1* and *EPHB1*, which was mutated in two patients with WBS.[36–39]

Epistatic effects of rare variants

To study if second-hit variants could be disrupting genes already altered in WBS and act by an epistatic effect, we intersected a manually curated list of genes with altered expression in WBS with the list of mutated genes (figure 3). The results showed that 20 of mutated genes were also altered in WBS. Two of them harboured a LoF in our cohort: *AGAP1* and *EPHB1* in two cases.

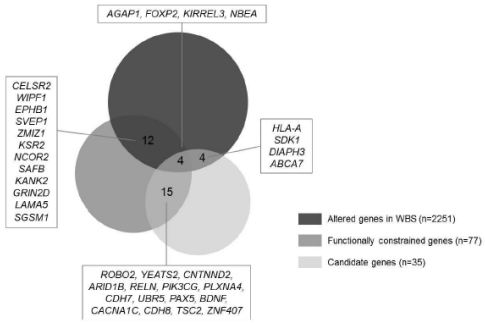


Figure 3 Overlap between candidate and functionally constrained genes mutated in our cohort and genes altered in Williams-Beuren syndrome (WBS).

Burden of rare deleterious mutations, increased in females

In total, six patients with WBS with ASD carried, in addition to the specific de novo 7q11.23 deletion, one to three strong candidate rare genetic variants (CNV and/or SNV) that could contribute to the ASD phenotype, either inherited or de novo (table 2).

Since ASD is 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.[40] 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 males (n=4) and females (n=4) (figure 4). The results showed a statistically significant increase (p=0.016) in the number of total mutations per patient in females ($\bar{x}$ = 6, SD=1.7) compared with males ($\bar{x}$ = 3, SD=1.2). This effect was not seen when comparing the burden of all rare variants between genders (p=0.51).

Association study with candidate ASD genes

To study the potential role of common variants, we analysed the loci previously associated with ASD and compared their variation frequency in our cohort with that of NFE ExAC population. A total of 645 single-nucleotide markers located in coding regions were analysed, of which only 13 obtained a significant p value (p<0.05) and none passed FDR correction. For 10 of the significantly over-represented SNVs in our cohort, the over-represented allele corresponded to the risk allele, whereas it was the protector allele for the 3 remaining. 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 with transcript RP1.130H16.16/*CCDC157*.[41] We also identified a nominally significant increase in the frequency of rs2135720, a non-synonymous SNP within *PCDH15* associated with lipid traits.[42]

Finally, we also genotyped two promoter variants at *SLC6A4* not captured in exome data that had been proposed as modifiers for the phenotypic outcome in WBS in two patients with autistic symptoms and hyperserotonaemia. 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 (*l*/*s*) and 1/8 individual being homozygous for the *s* allele.

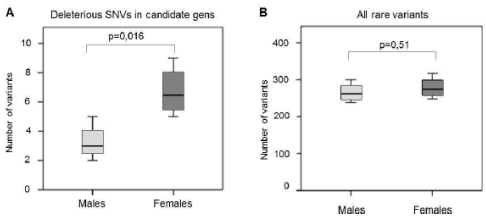


Figure 4 Gender (male vs female) comparison of the average number of rare variants in Williams-Beuren syndrome-autism spectrum disorders cases. Statistical significance in all tests was calculated using a two-sided Student's t-test, excluding SNVs of the X and Y chromosomes from the analysis. (A) Deleterious variants in candidate genes. (B) All rare variants. SNV, single nucleotide variant.

DISCUSSION

We have performed a genome-wide comprehensive analysis to investigate possible genetic factors in eight individuals presenting with both diagnoses of WBS and ASD. To date, the analysis of exome data failed to reveal genetic variants that could explain the variance of the social behaviour phenotype in a cohort of 85 patients with WBS without comorbid ASD,[43] and the only locus suggested to act as a modifying factor was the serotonin transporter *SLC6A4* (5-HTTLPR), based on two individuals with WBS and ASD who had hyperserotonaemia and were homozygous for the *ss* polymorphism at the gene promoter.[17] However, the results in our cohort show genotype frequencies similar to those of the general population and do not support a role of *SLC6A4*.

The characterisation of the deletion showed no atypical rearrangements and different breakpoints in the patients, excluding the alteration of flanking genes as a major cause. In the studied patients the deletion had originated in the paternal allele in seven of eight, suggesting that epigenetic control mechanisms could influence ASD risk in WBS. This parental bias was not present in our larger cohort of WBS without comorbid ASD (n=374 trios) as the origin of the deletion was 50% maternal and 50% paternal. There are some evidences supporting the imprinting of the 7q11.23 region based on expression of *FKBP6* and *GTF2I*.[44–46] However, in a recent study of six patients with WBS and ASD, the deletion was of maternal origin in four out of six patients.[47] Collectively, there is suggestive evidence regarding a possible role of genetic imprinting, but not enough data to confirm it, which demands characterisation of the deletion and parental origin in additional individuals with WBS and ASD to confirm whether imprinting can influence phenotypic variability.

Besides deletion origin and breakpoints, hemizygous variants unmasked in the deleted region could also act as second-hit modifying factors. The detailed analysis of SNVs did not reveal any common haplotype or any rare variant in the coding regions of *GTF2I* or *GTF2IRD2*. However, the analysis of common variants revealed an over-representation of a relatively rare hemizygous variant previously associated with ASD (rs12539160).[28] It has a MAF of 0.06 in European population and of 0.03 in the Spanish population but it was found in two patients (0.25) of our cohort, representing an eightfold increase in frequency.[32] While not meeting strict criteria for statistical significance (p=0.076), this finding deserves additional investigation in future studies. In a recent study, exome data of 85 patients with WBS was analysed in order to search for common and rare variants that could

account for the variance in the social behaviour of these patients.[43] Together with the genotyping of our cohort of non-autistic WBS, this variant (rs12539160) was found with a frequency of 0.0573 among 157 WBS individuals which is similar to general population. Although this variant lies in the intronic region of *MLXIPL*, mainly involved in lipid abnormalities, deregulation of the same or other nearby gene either by this SNP or other risk variants in linkage disequilibrium not detected by exome sequencing might contribute to ASD risk in WBS.

We also studied genetic variation at a genome-wide level. Regarding CNVs, no additional large (>500 kb) or pathogenic rearrangement was identified. Likewise, previous studies evaluating second-hit CNVs in genomic disorders showed that additional rearrangements were more frequent in disorders of variable expressivity than in syndromic entities. In fact, only 5% of individuals with WBS carried a second event, similar frequency than in the control population, suggesting that additional large CNVs may cause a more severe phenotype and/or be incompatible with life.[48] However, relatively small CNVs altering ASD candidate genes and therefore potential contributors to the ASD phenotype were found in two patients, a partial duplication of the last two exons of *SIK1*, and a complete heterozygous deletion of *DUSP22*.

The analysis of rare deleterious SNVs on ASD candidate and functionally constrained genes uncovered several brain-expressed genes that harbour LoF variants in our cohort. Among those, we detected a de novo stop mutation affecting *UBR5*, an E3 ubiquitin-protein ligase in a severely affected female. Therefore, the presence of WBS and ASD in this individual is probably due to the co-occurrence of two mechanistically unrelated de novo events, the WBS deletion and the *UBR5* mutation.

However, the increased prevalence of ASD features in WBS and the low frequency of de novo events suggest that other genetic factors of smaller effect may be influencing the ASD risk in most cases. In fact, we have identified several rare variants in ASD candidate genes. Two of the genes harbouring LoF mutations, *EPHB1* and *AGAP1*, have been found deregulated in WBS, suggesting a possible epistatic effect. *AGAP1* was hypermethylated in patients with WBS compared with individuals with the reciprocal duplication, which is associated with a phenotype of language impairment, anxiety and increased risk of ASD and schizophrenia.[24 49–51] *EPHB1* was underexpressed in patients with WBS and individuals with atypical deletions and low IQ. Inherited mutations in those genes would result in further deregulation of the expression already altered by the WBS deletion[21] and higher risk for a more severe phenotype.

Globally, six of the eight patients with WBS with ASD carried, in addition to the specific de novo deletion at the 7q11.23 locus, one to three rare genetic variants (CNV and/or SNV) altering candidate genes that could contribute to the ASD phenotype. Consequently, a proportion of cases of WBS with associated ASD could be explained by the same factors influencing ASD risk in the general population with a threshold model. Given the increased risk for ASD in WBS, the WBS deletion would act as a predisposing factor facilitating the role of other genetic (inherited or de novo) and/or environmental factors by acting on an already sensitised background.

Interestingly, the average number of rare deleterious variants in candidate genes was significantly higher in females than males, an effect not observed when considering all rare variants. Although our small sample size requires caution interpreting the results, this difference could support the higher risk threshold requiring higher genetic load in females to develop ASD, or could just be explained by the fact that females in our cohort had a more severe presentation than males. Further studies assessing the difference in prevalence and severity of ASD features between genders in individuals with WBS will help clarify if the female protective effect has a role in the expressivity of disorders of full penetrance.

Finally, we assessed the contribution of common variation by looking for over-represented variants previously associated with ASD. Although none of the variants passed FDR correction, three markers with nominal p values had been previously described as eQTLs in brain and may have a direct functional effect. However, our study is limited by the fact that most common variation is not covered by exome sequencing, as it resides in non-coding regions. Currently, common variation seems to explain at least 20% of ASD liability.[52] Therefore, further studies regarding the role of common variants in autism will provide a necessary basis for future studies.

In summary, our work represents a thorough assessment of second-hit modifier genetic factors in individuals with WBS and associated ASD. Similar to previous reports, patients did not differ in deletion breakpoints, discarding the role of atypical rearrangements. However, in seven of eight the deletion had originated in the paternal allele, a factor that had not been addressed in previous studies. In addition, two individuals carried a relatively rare hemizygous variant previously associated with ASD, representing an eightfold increase with respect to the general population. Inherited rare variants and a de novo nonsense mutation in ASD-related or functionally constrained genes were identified in six cases, with higher burden in females compared with males. Taken together, these results suggest that imprinting mechanisms, *trans*-acting factors in the remaining allele and inherited or de novo rare sequence variants elsewhere in the genome may play a role in the susceptibility to ASD in WBS. Therefore, similar factors influencing ASD risk in the general population also shape phenotypic variability in disorders with full penetrance.

Author affiliations
[1]Genetics Unit, Universitat Pompeu Fabra, Hospital del Mar Research Institute (IMIM), and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Universitat Pompeu Fabra Departament de Ciences Experimentals i de la Salut, Barcelona, Spain
[2]Clinical and Molecular Genetics Area, Vall Hebrón Hospital Research Institute (VHIR), Hospital Vall d'Hebron, Barcelona, Spain
[3]Fundacio Dexeus Salut de la Dona, Barcelona, Spain
[4]SA Clinical Genetics, Women's and Children's Hospital, South Australian Health and Medical Research Institute (SAHMRI) and University of Adelaide, South Australian Health and Medical Research Institute, Adelaide, South Australia, Australia

Acknowledgements The authors would like to thank the patients and their families for their support. The authors would also like to thank Marcos López-Sánchez for his help with the data acquisition and handling.

Contributors MC-S and MC-R analysed the exome data, performed the experiments and drafted the manuscript. DP-G and MGP-V contributed with the clinical and phenotypic information. RF contributed with detailed molecular characterisation of deletions. IC and LAP-J conceived the study and participated in the design and data interpretation, and helped in drafting the manuscript. All authors read and approved the manuscript.

Funding This work was funded by grants from the Spanish Ministry of Economy and Competiveness (FIS PI16/00369 and PI1302481 cofunded by FEDER, and 'Programa de Excelencia María de Maeztu' MDM-2014-0370), and the Generalitat de Catalunya (2017SRG01974 and ICREA-Acadèmia program). MC-R had a predoctoral fellowship of Ministry of Education, Culture and Sport (FPU16/06907) and DP-G had a predoctoral fellowship from the Instituto de Salud Carlos III (FI11/00656).

Competing interests LAP-J is scientific advisor of qGenomics SL. The remaining authors declare no conflict of interest.

Patient consent for publication Not required.

Ethics approval The study was approved by the Clinical Research Ethics Committee of the Parc Salut Mar.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as supplementary information.

REFERENCES

1 Strømme P, Bjørnstad PG, Ramstad K. Prevalence estimation of Williams syndrome. *J Child Neurol* 2002;17:269–71.
2 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;69:1–13.
3 Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome deletions. *Am J Hum Genet* 2003;73:131–51.
4 Howald C, Merla G, Digilio MC, Amenta S, Lyle R, Deutsch S, Choudhury U, Bottani A, Antonarakis SE, Fryssira H, Dallapiccola B, Reymond A. Two high throughput technologies to detect segmental aneuploidies identify new Williams-Beuren syndrome patients with atypical deletions. *J Med Genet* 2006;43:266–73.
5 Palacios-Verdú MG, Segura-Puimedon M, Borralleras C, Flores R, Del Campo M, Campuzano V, Pérez-Jurado LA. Metabolic abnormalities in Williams-Beuren syndrome. *J Med Genet* 2015;52:248–55.
6 Pober BR. Williams-Beuren syndrome. *N Engl J Med* 2010;362:239–52.
7 Merla G, Brunetti-Pierri N, Micale L, Fusco C. Copy number variants at Williams-Beuren syndrome 7q11.23 region. *Hum Genet* 2010;128:3–26.
8 Jones W, Bellugi U, Lai Z, Chiles M, Reilly J, Lincoln A, Adolphs R. II. Hypersociability in Williams syndrome. *J Cogn Neurosci* 2000;12 Suppl 1:30–46.
9 Asada K, Itakura S. Social phenotypes of autism spectrum disorders and Williams syndrome: similarities and differences. *Front Psychol* 2012;3:247.
10 Tordjman S, Anderson GM, Botbol M, Toutain A, Sarda P, Carlier M, Saugier-Veber P, Baumann C, Cohen D, Lagneaux C, Tabet A-C, Verloes A. Autistic disorder in patients with Williams-Beuren syndrome: a reconsideration of the Williams-Beuren syndrome phenotype. *PLoS One* 2012;7:e30778.
11 Gosch A, Pankau R. "Autistic" behavior in two children with Williams-Beuren syndrome. *Am J Med Genet* 1994;53:83–4.
12 Herguner S, Mukaddes NM. Autism and Williams syndrome: a case report. *World J Biol Psychiatry* 2006;7:186–8.
13 Reiss AL, Feinstein C, Rosenbaum KN, Borengasser-Caruso MA, Goldsmith BM. Autism associated with Williams syndrome. *J Pediatr* 1985;106:247–9.
14 August GJ, Realmuto GM. Williams syndrome: serotonin's association with developmental disabilities. *J Autism Dev Disord* 1989;19:137–41.
15 Gillberg C, Rasmussen P. Brief report: four case histories and a literature review of Williams syndrome and autistic behavior. *J Autism Dev Disord* 1994;24:381–93.
16 Klein-Tasman BP, van der Fluit F, Mervis CB. Autism spectrum symptomatology in children with Williams syndrome who have phrase speech or fluent language. *J Autism Dev Disord* 2018;48:3037–50.
17 Tordjman S, Anderson GM, Cohen D, Kermarrec S, Carlier M, Touitou Y, Saugier-Veber P, Lagneaux C, Chevreuil C, Verloes A. Presence of autism, hyperserotonemia, and severe expressive language impairment in Williams-Beuren syndrome. *Mol Autism* 2013;4:29.
18 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;13:444–67.
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. *Lancet Psychiatry* 2015;2:909–16.
20 Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, Van Vooren S, Moreau Y, Pettett RM, Carter NP. Decipher: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am J Hum Genet* 2009;84:524–33.
21 Antonell A, Del Campo M, Magano LF, Kaufmann L, de la Iglesia JM, Gallastegui F, Flores R, Schweigmann U, Fauth C, Kotzot D, Pérez-Jurado LA. 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;47:312–20.
22 Henrichsen CN, Csárdi G, Zabot M-T, Fusco C, Bergmann S, Merla G, Reymond A. Using transcription modules to identify expression clusters perturbed in Williams-Beuren syndrome. *PLoS Comput Biol* 2011;7:e1001054.
23 Adamo A, Atashpaz S, Germain P-L, Zanella M, D'Agostino G, Albertin V, Chenoweth J, Micale L, Fusco C, Unger C, Augello B, Palumbo O, Hamilton B, Carella M, Donti E, Pruneri G, Selicorni A, Biamino E, Prontera P, McKay R, Merla G, Testa G. 7Q11.23 dosage-dependent dysregulation in human pluripotent stem cells affects transcriptional programs in disease-relevant lineages. *Nat Genet* 2015;47:132–41.
24 Strong E, Butcher DT, Singhania R, Mervis CB, Morris CA, De Carvalho D, Weksberg R, Osborne LR. Symmetrical dose-dependent DNA-methylation profiles in children with deletion or duplication of 7q11.23. *Am J Hum Genet* 2015;97:216–27.
25 Eicher JD, Landowski C, Stackhouse B, Sloan A, Chen W, Jensen N, Lien J-P, Leslie R, Johnson AD. Grasp v2.0: an update on the genome-wide Repository of associations between SNPs and phenotypes. *Nucleic Acids Res* 2015;43:D799–D804.
26 Bass J, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control, 2015. Available: http://github.com/jdstorey/qvalue
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:e47457.
28 Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Sykes N, Pagnamenta AT, Almeida J, Bacchelli E, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Carson AR, Casallo G, Casey J, Chu SH, Cochrane L, Corsello C, Crawford EL, Crossett A, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu X-Q, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Melhem NM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Piven J, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Wing K, Wittemeyer K, Wood S, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Betancur C, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Gallagher L, Geschwind DH, Gill M, Haines JL, Miller J, Monaco AP, Nurnberger JI, Paterson AD, Pericak-Vance MA, Schellenberg GD, Scherer SW, Sutcliffe JS, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Devlin B, Ennis S, Hallmayer J. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 2010;19:4072–82.
29 Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niyazov D, Leppig K, Thiese H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE. A copy number variation morbidity map of developmental delay. *Nat Genet* 2011;43:838–46.
30 Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A. SFARI gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). *Mol Autism* 2013;4:36.
31 Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee Y-H, Yamrom B, Wigler M. Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci U S A* 2015;112:E5600–E5607.
32 Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91.
33 Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, Raja A, Coe BP, Stessman HA, He Z-X, Leal SM, Bernier R, Eichler EE. Excess of rare, inherited truncating mutations in autism. *Nat Genet* 2015;47:582–8.
34 De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, Kosmicki J, Shih-Chen F, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, Carracedo A, Chahrour MH, Chiocchetti AG, Coon H, Crawford EL, Curran SR, Dawson G, Duketis E, Fernandez BA, Gallagher L, Geller E, Guter SJ, Hill RS, Ionita-Laza J, Jimenz Gonzalez P, Kilpinen H, Klauck SM, Kolevzon A, Lee I, Lei I, Lei J, Lehtimäki T, Lin C-F, Ma'ayan A, Marshall CR, McInnes AL, Neale B, Owen MJ, Ozaki N, Parellada M, Parr JR, Purcell S, Puura K, Rajagopalan D, Rehnström K, Reichenberg A, Sabo A, Sachse M, Sanders SJ, Schafer C, Schulte-Rüther M, Skuse D, Stevens C, Szatmari P, Tammimies K, Valladares O, Voran A, Li-San W, Weiss LA, Willsey AJ, Yu TW, Yuen RKC, Cook EH, Freitag CM, Gill M, Hultman CM, Lehner T, Palotie A, Schellenberg GD, Sklar P, State MW, Sutcliffe JS, Walsh CA, Scherer SW, Zwick ME, Barett JC, Cutler DJ, Roeder K, Devlin B, Daly MJ, Buxbaum JD, Ercument Cicek A, Fu S-C, Crooks L, Sean Hill R, Ma'ayan A, Wang L-S, Jeremy Willsey A, S-C F, TW Y, Devlin B, Daly MJ, DDD Study, Homozygosity Mapping Collaborative for

Autism, UK10K Consortium. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014;515:209–15.
35 Lim ET, Uddin M, De Rubeis S, Chan Y, Kamumbu AS, Zhang X, D'Gama AM, Kim SN, Hill RS, Goldberg AP, Poultney C, Minshew NJ, Kushima I, Aleksic B, Ozaki N, Parellada M, Arango C, Penzol MJ, Carracedo A, Kolevzon A, Hultman CM, Weiss LA, Fromer M, Chiocchetti AG, Freitag CM, Church GM, Scherer SW, Buxbaum JD, Walsh CA, Autism Sequencing Consortium. Rates, distribution and implications of postzygotic mosaic mutations in autism spectrum disorder. *Nat Neurosci* 2017;20:1217–24.
36 Tate CM, Lee J-H, Skalnik DG. CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin. *Febs J* 2010;277:210–23.
37 Montgomery TR, Steinkellner T, Sucic S, Koban F, Schüchner S, Ogris E, Sitte HH, Freissmuth M. 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;34:6344–51.
38 Egea J, Klein R. Bidirectional Eph-ephrin signaling during axon guidance. *Trends Cell Biol* 2007;17:230–8.
39 Ikeda M, Aleksic B, Kinoshita Y, Okochi T, Kawashima K, Kushima I, Ito Y, Nakamura Y, Kishi T, Okumura T, Fukuo Y, Williams HJ, Hamshere ML, Ivanov D, Inada T, Suzuki M, Hashimoto R, Ujike H, Takeda M, Craddock N, Kaibuchi K, Owen MJ, Ozaki N, O'Donovan MC, Iwata N. Genome-Wide association study of schizophrenia in a Japanese population. *Biol Psychiatry* 2011;69:472–8.
40 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;110:5258–62.
41 Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, Foster B, Moser M, Karasik E, Gillard B, Ramsey K, Sullivan S, Bridge J, Magazine H, Syron J, Fleming J, Siminoff L, Traino H, Mosavel M, Barker L, Jewell S, Rohrer D, Maxim D, Filkins D, Harbach P, Cortadillo E, Berghuis B, Turner L, Hudson E, Feenstra K, Sobin L, Robb J, Branton P, Korzeniewski G, Shive C, Tabor D, Qi L, Groch K, Nampally S, Buia S, Zimmerman A, Smith A, Burges R, Robinson K, Valentino K, Bradbury D, Cosentino M, Diaz-Mayoral N, Kennedy M, Engel T, Williams P, Erickson K, Ardlie K, Winckler W, Getz G, DeLuca D, MacArthur D, Kellis M, Thomson A, Young T, Gelfand E, Donovan M, Meng Y, Grant G, Mash D, Marcus Y, Basile M, Liu J, Zhu J, Tu Z, Cox NJ, Nicolae DL, Gamazon ER, Im HK, Konkashbaev A, Pritchard J, Stevens M, Flutre T, Wen X, Dermitzakis ET, Lappalainen T, Guigo R, Monlong J, Sammeth M, Koller D, Battle A, Mostafavi S, McCarthy M, Rivas M, Maller J, Rusyn I, Nobel A, Wright F, Shabalin A, Feolo M, Sharopova N, Sturcke A, Paschal J, Anderson JM, Wilder EL, Derr LK, Green ED, Struewing JP, Temple G, Volpi S, Boyer JT, Thomson EJ, Guyer MS, Ng C, Abdallah A, Colantuoni D, Insel TR, Koester SE, Little AR, Bender PK, Lehner T, Yao Y, Compton CC, Vaught JB, Sawyer S, Lockhart NC, Demchok J, Moore HF, GTEx Consortium. The Genotype-Tissue expression (GTEx) project. *Nat Genet* 2013;45:580–5.
42 Anney R, Klei L, Pinto D, Almeida J, Bacchelli E, Baird G, Bolshakova N, Bölte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Casey J, Conroy J, Correia C, Corsello C, Crawford EL, de Jonge M, Delorme R, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Gilbert J, Gillberg C, Glessner JT, Green A, Green J, Guter SJ, Heron EA, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Jacob S, Kenny GP, Kim C, Kolevzon A, Kustanovich V, Lajonchere CM, Lamb JA, Law-Smith M, Leboyer M, Le Couteur A, Leventhal BL, Liu X-Q, Lombard F, Lord C, Lotspeich L, Lund SC, Magalhaes TR, Mantoulan C, McDougle CJ, Melhem NM, Merikangas A, Minshew NJ, Mirza GK, Munson J, Noakes C, Nygren G, Papanikolaou K, Pagnamenta AT, Parrini B, Paton T, Pickles A, Posey DJ, Poustka F, Ragoussis J, Regan R, Roberts W, Roeder K, Roge B, Rutter ML, Schlitt S, Shah N, Sheffield VC, Soorya L, Sousa I, Stoppioni V, Sykes N, Tancredi R, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Vorstman JAS, Wallace S, Wing K, Wittemeyer K, Wood S, Zurawiecki D, Zwaigenbaum L, Bailey AJ, Battaglia A, Cantor RM, Coon H, Cuccaro ML, Dawson G, Ennis S, Freitag CM, Geschwind DH, Haines JL, Klauck SM, McMahon WM, Maestrini E, Miller J, Monaco AP, Nelson SF, Nurnberger JI, Oliveira G, Parr JR, Pericak-Vance MA, Piven J, Schellenberg GD, Scherer SW, Vicente AM, Wassink TH, Wijsman EM, Betancur C, Buxbaum JD, Cook EH, Gallagher L, Gill M, Hallmayer J, Paterson AD, Sutcliffe JS, Szatmari P, Vieland VJ, Hakonarson H, Devlin B. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet* 2012;21:4781–92.
43 Kopp ND, Parrish PCR, Lugo M, Dougherty JD, Kozel BA. Exome sequencing of 85 Williams-Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior. *Mol Genet Genomic Med* 2018;6:749–65.
44 Collette JC, Chen X-N, Mills DL, Galaburda AM, Reiss AL, Bellugi U, Korenberg JR. William's syndrome: gene expression is related to parental origin and regional coordinate control. *J Hum Genet* 2009;54:193–8.
45 Strogantsev R, Krueger F, Yamazawa K, Shi H, Gould P, Goldman-Roberts M, McEwen K, Sun B, Pedersen R, Ferguson-Smith AC. Allele-Specific binding of Zfp57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol* 2015;16:112.
46 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;59:781–92.
47 Masson J, Demily C, Chatron N, Labalme A, Rollat-Farnier P-A, Schluth-Bolard C, Gilbert-Dussardier B, Giuliano F, Touraine R, Tordjman S, Verloes A, Testa G, Sanlaville D, Edery P, Lesca G, Rossi M. Molecular investigation, using chromosomal microarray and whole exome sequencing, of six patients affected by Williams Beuren syndrome and autism spectrum disorder. *Orphanet J Rare Dis* 2019;14:121.
48 Girirajan S, Rosenfeld JA, Coe BP, Parikh S, Friedman N, Goldstein A, Filipink RA, McConnell JS, Angle B, Meschino WS, Nezarati MM, Asamoah A, Jackson KE, Gowans GC, Martin JA, Carmany EP, Stockton DW, Schnur RE, Penney LS, Martin DM, Raskin S, Leppig K, Thiese H, Smith R, Aberg E, Niyazov DM, Escobar LF, El-Khechen D, Johnson KD, Lebel RR, Siefkas K, Ball S, Shur N, McGuire M, Brasington CK, Spence JE, Martin LS, Clericuzio C, Ballif BC, Shaffer LG, Eichler EE. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med* 2012;367:1321–31.
49 Mulle JG, Pulver AE, McGrath JA, Wolyniec PS, Dodd AF, Cutler DJ, Sebat J, Malhotra D, Nestadt G, Conrad DF, Hurles M, Barnes CP, Ikeda M, Iwata N, Levinson DF, Gejman PV, Sanders AR, Duan J, Mitchell AA, Peter I, Sklar P, O'Dushlaine CT, Grozeva D, O'Donovan MC, Owen MJ, Hultman CM, Kähler AK, Sullivan PF, Kirov G, Warren ST, Molecular Genetics of Schizophrenia Consortium. Reciprocal duplication of the Williams-Beuren syndrome deletion on chromosome 7q11.23 is associated with schizophrenia. *Biol Psychiatry* 2014;75:371–7.
50 Deshpande A, Weiss LA. Recurrent reciprocal copy number variants: roles and rules in neurodevelopmental disorders. *Dev Neurobiol* 2018;78:519–30.
51 Mervis CB, Klein-Tasman BP, Huffman MJ, Velleman SL, Pitts CH, Henderson DR, Woodruff-Borden J, Morris CA, Osborne LR. Children with 7q11.23 duplication syndrome: psychological characteristics. *Am J Med Genet A* 2015;167:1436–50.
52 Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, Samocha KE, Goldstein JI, Okbay A, Bybjerg-Grauholm J, Werge T, Hougaard DM, Taylor J, Skuse D, Devlin B, Anney R, Sanders SJ, Bishop S, Mortensen PB, Børglum AD, Smith GD, Daly MJ, Robinson EB, iPSYCH-Broad Autism Group, Psychiatric Genomics Consortium Autism Group. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet* 2017;49:978–85.

SUPPLEMENTARY MATERIAL

The supplementary material is available at the Journal of Medical Genetics web page: https://jmg.bmj.com/content/56/12/801.long

It is not included in this thesis due to inappropriate format and length. The supplementary information is listed below:

Supplementary Table 1. List of deregulated genes in WBS obtained from literature (n=2,251).

Supplementary Table 2. List of SNPs previously associated with ASD located in exonic regions, according to GRASP.

Supplementary Table 3. List of candidate genes from SFARI (n=791).

Supplementary Table 4. List of all rare deleterious SNVs found in each patient including variants in candidate genes and highly constrained genes.

CHAPTER 2

Transcriptomic characterization of 7q11.23 patient-specific induced pluripotent stem cells (iPSCs) and derivatives

Mar Costa-Roger[1], Bernd Kuebler[2], Raquel Flores[1], Nerea Ugartondo[3], Ainhoa Arcas-Márquez[3], Marina Alvarez-Estape[1], Anna Veiga[2,4], Luis A. Pérez-Jurado[1,5], Ivon Cuscó[1,6]*, Roser Corominas[1,3]*

In preparation

ABSTRACT

Introduction. Williams-Beuren syndrome (WBS; OMIM #194050) and 7q11.23 microduplication syndrome (DUP7; OMIM #609757) are rare multisystem disorders with somehow opposed neurobehavioural trajectories caused by 1.55-1.83 Mb heterozygous microdeletion or microduplication of 28-30 contiguous genes at 7q11.23, respectively. Cellular reprogramming is a good approach to overcome the experimental limitations to study neurodevelopmental disorders in humans. The purpose of this project is to evaluate the transcriptomic consequences of 7q11.23 patient-derived iPSC lines and derivatives.

Methods. We generated patient-specific iPSCs from fibroblasts from four WBS patients, four Dup7 patients and two controls, which were differentiated to neural progenitor cells (NPCs) and to dopaminergic neurons (iNeus). After RNA extraction from all cell types (fibroblasts, iPSCs, NPCs and iNeus), we assessed genome-wide differential expression using expression microarrays.

Results. We identified a total of 494 unique differentially expressed genes (DEGs) in a pairwise comparison of the three genotypes in all cell types. Half of the genes in 7q11.23 showed mirroring expression between Dup7 and WBS models. Interestingly, *MIR590* presented the highest log_2FC values and its predicted targets were enriched in ASD risk genes and genes previously related to WBS by transcriptomic studies. Enrichment analysis of DEGs in fibroblasts revealed specific pathways and gene ontology categories relevant for the hallmark phenotypes of both disorders. Neuronal processes, such as transmission across chemical synapses, were significantly enriched in neurons. In addition, WGCNA of iPSCs and NPCs lines

uncovered interesting co-expression modules related to cholesterol biosynthesis (q-value=4.44E-5) or glutamate receptor signaling (q-value=0.0463), respectively.
Conclusions. We present one of the largest 7q11.23 patient-derived iPSC lines and derivatives including both duplication and deletion models. Gene expression from 7q11.23 region has proved to be consistent, further supporting expression imbalances of *BAZ1B, BCL7B, TBL2, WBSCR22, EIF4H, RFC2, CLIP2* and *GTF2IRD1*, and highlighting *MIR590* as a novel compelling candidate to regulate the expression of genes relevant for both Dup7 and WBS disorders. Integrative transcriptomic analysis of *in vitro* 7q11.23-CNVs cellular models revealed candidate genes and pathways related to cholesterol metabolism, Arf6 trafficking events and glutamate signaling altered during early neuronal development in these genomic disorders, which could lead to novel potential therapeutic targets.

Keywords: 7q11.23 microduplication syndrome; Williams-Beuren syndrome; iPSC.

[1]Genetics Unit, Universitat Pompeu Fabra, Hospital del Mar Research Institute (IMIM), Parc de Salut Mar and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Universitat Pompeu Fabra, Departament de Ciències Experimentals i de la Salut, Barcelona, Spain

[2]Regenerative Medicine Programme, Institut d'Investigació Biomèdica de Bellvitge, IDIBELL, L'Hospital de Llobregat, and Spanish National Stem Cell Bank-Barcelona Node, Instituto de Salud Carlos III (ISCIII), Madrid, Spain

[3]Department of Genetics, Microbiology and Statistics, Barcelona University, Barcelona, Spain

[4]Reproductive Medicine Service, Dexeus Mujer, Hospital Universitari Dexeus, Barcelona, Spain

[5]Genetics Service, Parc de Salut Mar, Barcelona, Spain

[6]Department of Clinical and Molecular Genetics and Medicine Genetics Group, VHIR, University Hospital Vall d'Hebron, Barcelona, Spain

INTRODUCTION

7q11.23 *locus* is a complex genomic region bearing rare recurrent rearrangements (1). Reciprocal copy number variants (CNVs) lead to a paradigmatic pair of neurodevelopmental disorders. Williams-Beuren syndrome (WBS; OMIM #194050) and 7q11.23 microduplication syndrome (Dup7; OMIM #609757) are rare multisystem disorders caused by a 1.55-1.83 Mb heterozygous microdeletion or microduplication of ~30 contiguous genes, respectively (2). This genomic region is composed of a single-copy genomic segment of ~1.2 Mb flanked by three large groups of segmental duplications or low copy repeats (LCR) (centromeric, medial and telomeric) with three differentiated blocks A, B and C each. The LCRs high homology can cause a misalignment in meiosis leading to non-allelic homologous recombination (NAHR) events generating gametes carrying the 7q11.23 deletion or duplication (3–5).

WBS is estimated to affect ~1/7500 live births (6). Patients are characterized by distinctive facial features, cardiovascular abnormalities (mainly supravalvular aortic stenosis), connective tissue alterations, growth and endocrine defects and a unique cognitive and behavioural profile (7). In particular, the neurobehavioural hallmark combines mild to moderate intellectual disability (ID) and severe impairment of visuospatial construction together with hipersociability and relatively well-preserved language abilities (7,8). The 7q11.23 microduplication syndrome affects approximately ~1/7500-20000 individuals (9,10). Dup7 individuals show significant speech and language delay, mild facial dysmorphism, cardiovascular problems (mostly aortic dilation), growth and development anomalies, and is frequently associated with autism spectrum disorders (ASD) or schizophrenia (11–15). Although WBS and Dup7 are somehow considered symmetrically opposite disorders, they also share common features such as ID, anxiety and ASD (2).

The molecular mechanisms and genotype-phenotype correlation of the genes in the 7q11.23 interval have been the focus of multiple investigations (16–18).

In addition, atypical rearrangements with breakpoints outside the LCRs have also participated in the establishment of these associations (19). It has been consistently demonstrated that the deletion of *ELN* (elastin) is responsible for the cardiovascular phenotype, while genes from the *GTF2I* (general transcription factor IIi) family are involved in cognitive and social phenotype (19,20).

Genome-wide expression studies from WBS patients' samples from different tissues have been used to identify genes with altered expression profiles. Expression microarrays of WBS lymphoblastoid cell lines discovered deregulated genes involved in the glycolysis pathway, neuronal migration and microtubule formation as well as genes potentially involved in the visuospatial deficits, highlighting *GTF2I, GTF2IRD1* (GTF2I repeat domain containing 1), *LIMK1* (LIM domain kinase 1) and *MAP1B* (microtubule associated protein 1B) (21). Transcriptomic profiling of skin fibroblasts from WBS patients detected genes significantly enriched in extracellular matrix, major histocompatibility complex and products of postsynaptic membrane (22). Lastly, microarray analysis and RNA sequencing (RNAseq) of peripheral blood from WBS individuals uncovered four co-expression modules associated with the neuropsychiatric phenotype of the disorder, and also found significantly deregulated genes related to language development and evolution (23,24). The most suitable accessible tissue to perform these studies is still debatable.

Given the difficulty to study functional defects in the developing human brain, the generation of *in vitro* cellular models is a good strategy to shed light on the pathophysiology of these disorders. The advent of cellular reprogramming techniques such as induced Pluripotent Stem Cells (iPSC)-based technology has allowed overcoming several limitations arisen from animal models and human studies such as the sample origin (25). Previous iPSC-based expression studies for 7q11.23 disorders have revealed diverse disease-relevant pathways (26–29). The first transcriptomic study of 7q11.23 CNVs in patient's derived iPSCs identified 757 differentially expressed genes (DEGs) in the pluripotent state with significant Gene Ontology (GO) enrichments for biological processes relevant

to the disease phenotypes such as cell adhesion, migration and motility, nervous system, inner ear morphogenesis and blood vessel development. For the differentiated neural progenitors, DEGs were mainly enriched for neuronal function categories (26). Additional studies have focused on gene expression arrays in iPSC-derived neurons from skin fibroblasts of WBS patients unravelling a significant transcriptional deregulation in neurotransmitter receptor activity, synapse assembly and synaptic transmission, and cognition (27,28). RNAseq of iPSC-derived neural progenitors and neurons from dental pulp cells of WBS subjects also revealed disease-relevant GO categories (29). These studies included between two and four different WBS patients' derived iPSCs, but only three patients with Dup7 (from Adamo and Zanella *et al.*) (26,30). In consequence, replications are needed in independent cohorts of patients with the 7q11.23 deletions or duplications.

In the present work, we have studied one of the largest cohorts of patient-specific iPSC lines and derivatives up to date, including four patients with WBS, four patients with Dup7, and two controls. We assessed genome-wide differential expression profiles in fibroblasts, iPSC lines, NPCs and differentiated neurons by microarray. In addition, we carried out an integrative transcriptomic analysis including enrichment analysis (pathways, GO terms and lists of genes associated to relevant clinical phenotypes) and weighted gene co-expression network analysis (WGCNA). A flowchart illustration of the strategy followed is depicted in Figure 1. Our results replicate previous independent studies and highlight novel players to be involved in 7q11.23 CNVs pathophysiology.

METHODS

Sample collection

We included four patients with WBS (7q11.23 deletion) and four patients with Dup7 (7q11.23 duplication), comprising two females and two males in each group. All individuals were already exhaustively characterized at molecular level, including analysis of CGH/SNP arrays, MLPA (multiplex ligation-dependent probe amplification) and microsatellite genotyping (Supplementary Table 1).

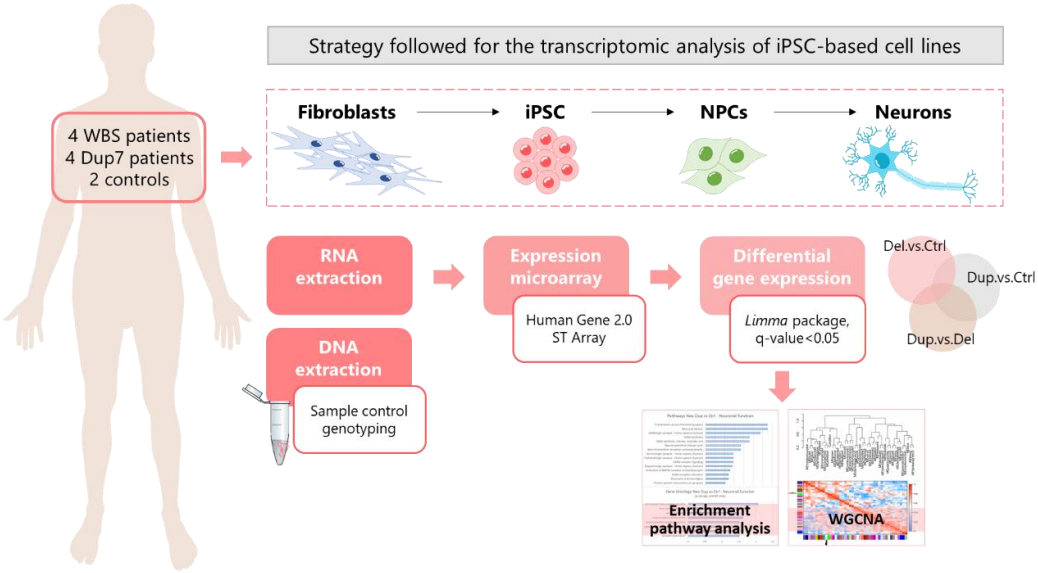


Figure 1. Flowchart illustrating the strategy followed for the integrative transcriptomic analysis. WBS: Williams-Beuren syndrome, Dup7: 7q11.23 microduplication syndrome, iPSCs: induced pluripotent stem cells, NPCs: neural progenitor cells, WGCNA: Weighted gene co-expression network analysis.

Written informed consent from parents or legal caregivers was obtained in order to perform skin biopsies. This project has been approved by the PRBB ethical committee, the Comité de Ética e Investigación Clinica-CEIC-CMRB and by the Catalan Authority for Stem Cell Research (Approval numbers: 05/2011, 3/2015, 8/2016).

iPSC generation

We generated patient-specific human iPSCs from fibroblasts of a total of 8 patients including at least two clones per individual. Additionally, we obtained iPSCs from 2 healthy controls (without CNV rearrangements at 7q11.23) from the Banco Nacional de Líneas Celulares (http://www.isciii.es), with 1-2 clones available. The iPSC generation from fibroblast using retroviral integrative vectors and their characterization has been thoroughly described in Kuebler B, et al. (31,32).

NPC induction

Patient-specific human iPSCs were differentiated to neural progenitor cells (NPCs) in feeder-free Matrigel® plates (protein concentration 0.25mg/mL, Corning), following a Gibco (Thermo Fisher Scientific) standard protocol (33), which is based on the use of neural induction medium avoiding the arduous procedure of mechanical NPC isolation from embryoid bodies. In brief, NPCs induction is stimulated during 7 days with a serum-free neural induction medium (Neurobasal Medium 98% and Neural Induction Supplement 2%) followed by harvest and expansion of NPCs for ~5 passages (~25 days, 4-6 days/passage) in maturation medium (Neurobasal Medium 49%, Neural Induction Supplement 2%, and Advanced DMEM/F-12 49%, all from Gibco) (Supplementary Figure 1A). Cells were passaged with StemPro® Accutase cell dissociation Reagent (Gibco), resuspended with scrapper and micropipette, and filtered by a 100 µM strainer before seeding at a density of 5-10x10^5 cells/cm^2.

In order to confirm the NPCs identity, we performed gene specific expression studies using quantitative RT-PCR and immunostaining followed by imaging (Supplementary Figure 2). The qRT-PCR was performed with Power SYBR Green PCR Master Mix (Roche) in an ABIPrism 7900 thermocycler (Applied Biosystems). mRNA expression levels of endogenous markers *OCT4* (specific from iPSCs), *PAX6* and *SOX1* (specifics from NPCs) were studied. Ct values were normalized using *GADPH* as a housekeeping gene and data analyzed with 2ΔCt method. Primers are listed on Supplementary Table 2.

For immunocytochemistry assays, cells were fixed with 4% paraformaldehyde in PBS washed with PBS 1x, incubated 1h in blocking solution (0.1% Triton-X100 in PBS) at room temperature (RT). Primary antibodies were diluted in blocking solution and incubated overnight at 4ºC in wet chambers [anti-MAP2 (1:25, mouse), anti-TUJ1 (1:50, mouse), anti-SYN1 (1:25, rabbit), anti-TH (1:100, rabbit), anti-GABA (1:50, rabbit anti-NESTIN (1:50, mouse), anti-SOX2 (1:50, rabbit), anti

FOXA2]. After washing, slides were incubated with secondary antibodies in 3% fetal bovine serum in PBS 1h at RT in the dark (Alexa Fluor 555/488 1:250). Hoechst33342 was used for DNA staining. For NPC characterization, we used Human Neural Stem Cell Immunocytochemistry Kit (Life technologies) that includes antibodies anti-NESTIN, anti-SOX1, anti-SOX2 and anti-PAX6. Fluorescence was detected using Olympus BX51 Fluorescence Microscope and confocal images were taken using Leica TSC SPE or Leica SP5 microscopes.

Dopaminergic neuron differentiation

We differentiated a subset of NPCs lines (n=10) towards dopaminergic neurons (iNeu) using a previously described protocol (33). Briefly, the derivation was carried out under feeder free conditions with specific growth factors in the medium during 20 consecutive days (Supplementary Figure 1B). Neuron generation was induced for the first 10 days using a specific medium containing Neurobasal Medium, MEM non-essential amino acids, GlutaMAX, B27 supplement, 200 ng/ml SHH and 100 ng/ml FGF8 (all from Gibco). For the last 10 days, the medium was composed of Neurobasal medium, MEM non-essential amino acids, GlutaMAX, B27 supplement, 200 µM ascorbic acid, 20 ng/ml BDNF and 20 ng/ml GDNF (all from Gibco). The efficacy of this protocol was evaluated by RT-qPCR (*NURR1, TH* and *FOXA2*; primers listed on Supplementary Table 2) and immunocytochemistry for dopaminergic neuron specific markers (TH, Tuj1 and FoxA2).

Acid nucleic extraction

DNA and total RNA were extracted from cell pellets using Gentra Puregene Kit (Qiagen) and RNeasy Mini Kit (Qiagen), respectively, following the manufacturer's protocol. Purity and integrity of the DNA and RNA were assessed by electrophoresis and spectrophotometry using the NanoDrop ND-2000 spectrophotometer (NanoDrop Technologies). In addition, RNA integrity and quality was assessed using a Bioanalyzer 2100 (Agilent Technologies). Only

samples with high purity and high integrity were subsequently used in microarray experiments (>8 RIN, except few samples).

Sample control genotyping

Line identity of all samples and cell types was corroborated by STRs (short tandem repeats) analysis of multiple microsatellite (D2S160, D5S406, D8S260, DS11S905, D13S158, D16S3091, D17S928, D20S171, D22S280), including two microsatellites at the 7q11.23 single-copy region (CR16t, WS5-ELN) (Supplementary Table 2).

Expression microarray

RNA samples were processed and analyzed at the Microarray Analysis Service (SAM) core facility from the Hospital del Mar Medical Research Institute (IMIM) using the Affymetrix GeneChip® platform. The manuals GeneChip WT PLUS Reagent kit (P/N 703174 Rev. 2) and Expression Wash, Stain and Scan User Manual (P/N 702731 Rev. 3) (ThermoFisher) were followed. Samples were hybridized in Affymetrix Human Gene 2.0 ST Array (ThermoFisher) in a GeneChip hybridization oven 640.

For statistical analysis, R program (Version 3.3.2) was used, together with different packages from Bioconductor (34) and the Comprehensive R Archive Network (CRAN) (35). Normalization was done using the Robust Multi-array Average algorithm (RMA) (36) included in *aroma.affymetrix* package (37).

Differential gene expression

To perform differential expression (DE) analysis, the *limma* package (38) was used treating the variable patient as a random effect in the model. Gene annotation was provided by Affymetrix for the HuGene-2_0-st array (NetAffx na36, hg19 genome assembly). The location of each Transcript Cluster ID (TCI) (Start, Stop, Strand and Chromosome) was used to map into the UCSC database

(Jun. 2013 hg19, GRCh37) to obtain genes matching the same coordinates. The tool liftOver in UCSC (39) was used to convert the hg19 version of genome assembly to the most recent hg38 version. TCIs that were not mapped to genes in the UCSC database or not belonging to one of the 24 human chromosomes were excluded from the analysis. Several pairwise comparisons were performed taking into account tissue type (Fibroblasts, iPSCs, NPCs and iNeu) and genotype categories (7q11.23 deletion, 7q11.23 duplication and control): Del.*vs*.Ctrl, Dup.*vs*.Ctrl and Dup.*vs*.Del. Genes were considered to be differentially expressed genes (DEGs) when the q-value was lower than 0.05 (5% False Discovery Rate, FDR).

In order to identify the DEGs overlapping between tissues and genotype comparisons, Venn's diagrams were constructed with two web tools, BioVenn (https://www.biovenn.nl) (40) and Venn diagrams from Bioinformatics & Evolutionary Genomics (http://bioinformatics.psb.ugent.be/webtools/Venn/).

Enrichment analysis

To elucidate possible deregulated mechanisms, pathway (KEGG, Reactome, etc.) and Gene Ontology (GO) enrichment analyses were performed using ConsensusPathDB database (CPDB, http://cpdb.molgen.mpg.de/) (41), a publicly available database that integrates interaction networks from 32 public resources. Statistical analyses were done with the CPBD overrepresentation analysis including pathway-based sets and GO categories (biological process and molecular function). All the human expressed genes were used as the background gene list. The adjusted p-value (q-value) threshold of 0.05 was used (5% FDR). A second approach to identify enriched pathways and GO categories using the web-based portal Metascape (https://metascape.org) (42) was performed obtaining similar results. For interpretation and representation of the results, we prioritized pathways categories over GO terms, with special interest in KEGG and Reactome databases.

Gene set over-representation analysis

We performed enrichment analysis of lists of ASD and WBS associated genes. 992 ASD risk genes were compiled from Simons Foundation Autism Research Initiative (SFARI) database (as of October 2020) (43) and the list of 2251 genes associated with WBS was extracted from Codina-Sola et al. (44), which included altered WBS genes from transcriptional deregulation, differential methylation and direct targets of *GTF2I*.

MicroRNA targets enrichment analysis

Given that we observed an altered expression in *MIR590*, we evaluated its possible implication in post-transcriptional regulation. We extracted a list of *MIR590* (hsa-miR-590) targets, separately for 5p and 3p, from the databases miRmap (45), miRWalk (46), miRTarBase (47) and miRsystem (48). We selected only those targets that were reported in at least two of them. We obtained a total of 1307 target genes for hsa-miR-590-5p and 3089 targets for has-miR-590-3p, 548 of them being targets of both microRNAs. We checked how many of those *MIR590* targets were differentially expressed in our analysis. Additionally, we intersected the list of targets with the lists of genes associated with ASD and WBS in order to perform enrichment analysis. We carried out a permutation test to evaluate the statistical significance of these results and to obtain empirical p-values. We randomly generated 10,000 gene sets of 1307 or 3089 genes from RefSeq genes and checked the number of genes present in the SFARI and WBS-related genes lists.

Weighted correlation network analysis

We performed weighted-gene co-expression network analysis (WGCNA) separately in iPSCs and NPCs using the WGCNA R Package following the approach described by Langfelder and Horvath (49). Data cleaning and pre-processing steps following the step-by-step gene network construction and module detection protocols lead to the removal of one outlier sample in each.

For iPSCs, clone 4F7 from control 1, and for NPCs clone 1.1 from SW112 patient. The soft-thresholding power β was evaluated with the *pickSoftThreshold* function of the WGCNA and was set to 4 to calculate co-expression similarity and adjacency matrices. Subsequently, the adjacency matrix was converted into a topological overlap matrix to detect gene connectivity. Module identification was performed with the dynamic tree cut method with a minimum module size of 30 and a deep split level of 2. Highly similar modules were merged with a height cut of 0.25. Lastly, we identified the modules significantly associated with the trait of interest (7q11.23 copies). Module-hub genes were described as genes in the module with |MM|>0.8 (Module Membership) and |GS|>0.3 (Gene Significance). GO and pathways enrichment analyses for the identified modules were performed as described before.

RESULTS

Establishment of iPSC-derived neuronal models from WBS, Dup7 and controls

We successfully generated NPCs from 20 iPSC clones derived from fibroblasts from four patients with WBS, four patients with Dup7 and two controls. After 5 passages of maturation, the generated NPC lines presented the appropriate morphology and were properly characterized by RT-qPCR and immunostaining of neuronal progenitor specific markers (NESTIN, SOX2, PAX6 and SOX1, Supplementary Figure 2).

The neuronal induction of iPSC lines from patient WBS159 was particularly problematic and required multiple attempts, probably due the fact that the patient WBS159, in addition to the 1.55Mb 7q11.23 deletion, carried a ~1.6 Mb microduplication at 16p13.11. This additional rare CNV was identified by SNP array (chr16:14,749,020-16,301,530 hg19) and confirmed by MLPA and

microsatellite genotyping (D16S3060). This microduplication involves around 30 genes and has been associated with psychiatric phenotypes (50).

Then, eight lines of NPCs from four patients (two deletions and two duplications) were differentiated to mature neurons (iNeu) following a dopaminergic neuron differentiation protocol (Supplementary Figure 1B). The morphological characteristics of the patients' lines apparently presented slower proliferation rate compared to controls and subtle variations that could not be quantified. All lines were able to establish neural networks. Immunocytochemistry assays confirmed the generation of neurons *(TUJ1+)* in all cases, but we observed differences in the number of *TH+* and *VGLUT+* (Supplementary Figure 3). The rest of the lines were not differentiated due to economic constraints and the observation of such heterogeneity.

To investigate global expression deregulation at the multiple *in vitro* differentiation stages (fibroblasts, iPSC, NPCs, iNeu), we extracted RNA and performed microarray expression analyses of all cell types.

Mirroring expression of 7q11.23 genes between WBS and Dup7 models

In order to ascertain the transcriptomic consequences of imbalances in 7q11.23 dosage, we analyzed the microarray data to identify DEGs. First, we focused on the expression pattern of the genes located in the 7q11.23 single-copy region in Dup7 (duplication model) and WBS (deletion model) compared to controls in all cell types. Specifically, 15 out of 30 genes (*NSUN5, BAZ1B, BCL7B, TBL2, DNAJC30, WBSCR22, STX1A, ABHD11, WBSCR27, LIMK1, EIF4H, MIR590, RFC2, GTF2IRD1* and *GTF2I*) presented a symmetrically opposite expression pattern with upregulation in Dup7 and downregulation in WBS compared to controls in all cell types (Figure 2). The other remaining genes were found upregulated or downregulated in the same direction in both models, which varied depending on the gene and cell type. For instance, *TRIM50* (tripartite motif containing 50)

was found downregulated both in Dup7 and WBS in all cell lines, except for neurons. Conversely, *GTF2IRD2* (GTF2I repeat domain containing 2) was upregulated for both duplication and deletion groups in fibroblasts and downregulated in the others. Therefore, we observed two differential tendencies regarding 7q11.23 gene expression between Dup7 and WBS: a symmetrically opposite expression or an alteration in the same direction. Despite the differential tendencies identified, only four genes (*BAZ1B, BCL7B, CLIP2* and *GTF2IRD1*) were significantly DE with a mirrored expression after adjusting statistical significance for multiple corrections (q-value<0.05, Figure 2). Our gene expression patterns are in line with previously reported transcriptional studies of iPSC-based cells (and derivatives), with matching results ranging between 92-96% depending on the cell type considered (22,26–28). Concretely, 11 genes (*NSUN5, BAZ1B, BCL7B, TBL2, DNAJC30, WBSCR22, EIF4H, RFC2, CLIP2, GTF2IRD1* and *GTF2I*) were consistently found upregulated in Dup7 and downregulated in WBS (Supplementary Table 3). In addition, the highest correlation was found in iPSC lines, with 18-20 genes out of 30 in the same direction.

MIR590 regulates the expression of genes enriched in ASD and WBS

Interestingly, we observed that *MIR590* differential expression presented the highest values in Dup7 and the lowest in WBS, suggesting that its expression is severely affected by 7q11.23 CNVs (Figure 2). Although the expression of *MIR590* in cases compared to controls did not reach adjusted statistical significance, when samples with 7q11.23 duplication were directly compared to samples with 7q11.23 deletion, *MIR590* was significantly altered in iPSCs ($log_2FC=2.07$, q-value=$3.90x10^{-5}$), NPCs ($log_2FC=1.98$, q-value=0.00012) and iNeu ($log_2FC=2.75$, q-value= 0.0018), suggesting a major role of this microRNA in early neural development.

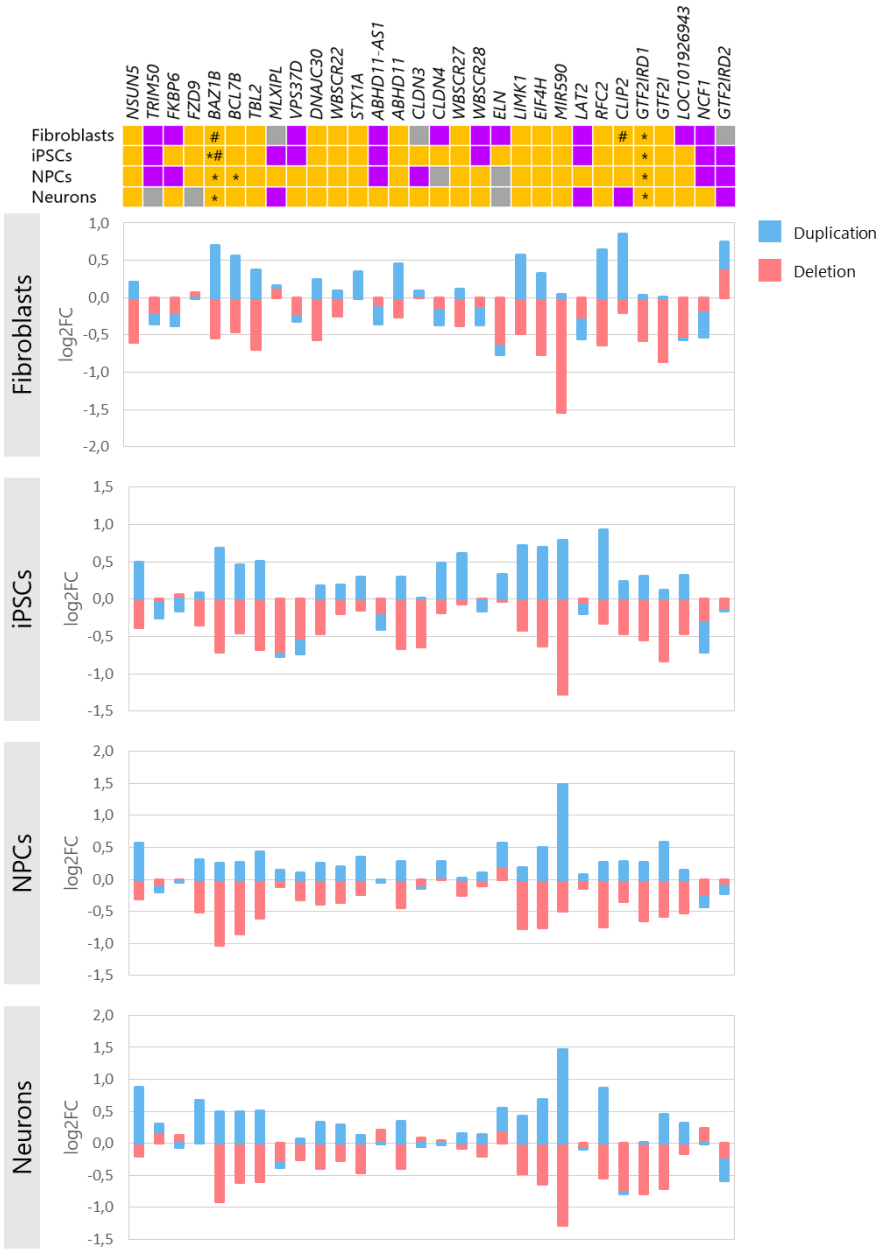


Figure 2. Gene expression pattern of genes in the 7q11.23 interval in the different *in vitro* models. Gene expression representation (log_2FC) for the comparisons: Deletion vs. Control (red) and Duplication vs. Controls (blue) in fibroblasts, iPSCs, NPCs and iNeu. On the top, schematic representation of the comparisons showing as yellow squares genes with opposite expression (mirrored) compared to controls. Grey squares indicate an upregulation in both WBS and Dup7 compared to controls, whereas violet squares denote a downregulation also in both categories. *: q-value < 0.05 in Deletion vs Controls, #: q-value < 0.05 in Duplication vs Controls.

To investigate the genes which post-transcriptional regulation is potentially regulated by *MIR590*, we compiled a list of *MIR590* targets identifying a total of 1307 genes regulated by hsa-miR-590-5p and 3089 targets by hsa-miR-590-3p, with 548 in common. Among the identified *MIR590* targets, we found several genes differentially expressed in our microarray analysis (52/1307 for 5p and 113/3089 for 3p in iNeu), but these numbers were not significant by permutation test. In addition, we intersected the list of *MIR590* targets with the lists of genes associated with ASD and WBS. Remarkably, the results showed that around 8% (106/1307 for 5p and 260/3089 for 3p) of *MIR590* targets were reported in SFARI, and approximately 12.5% (167/1307 for 5p and 388/3089 for 3p) were previously related to WBS. After random permutation test, we obtained empirical p-values<0.0001 for both miRNAs. Thus, MIR590 targets are enriched in ASD-related genes as well as genes linked to WBS.

Genome-wide transcriptional differences between WBS, Dup7 and controls

Next, we assessed the genome-wide differential expression considering exclusively genes presenting significant q-value<0.05. The four cell types were very clearly differentiated at expression level, and no batch effect was detected by combat (Supplementary Figure 4). A pairwise comparison of the three genotypes in all cell types (fibroblasts, iPSCs, NPCs and iNeu) identified a total of 494 unique DEGs (Supplementary Table 4). In the Dup.*vs*.Del comparison, the category with the highest statistical power, 18 genes overlapped at least in two cell type categories, from which 15 belonged to the 7q11.23 region, including six significantly deregulated in all cell types (*BAZ1B, EIF4H, BCL7B, LIMK1, GTF2IRD1* and *TBL2*) (Supplementary Figure 5). In addition, we highlight the genes *ZNF726*, *ZNF439* and *LOC105374122* (a suggested non-coding RNA), for being the only DEGs located outside the 7q11.23 interval deregulated in more than one cell type. *ZNF726* deregulation was common between iPSCs and NPCs, whereas *ZNF439* between NPCs and neurons.

<u>Fibroblasts</u>

The majority of DEGs were observed in fibroblasts. These genes included 218 Dup7 DEGs (Dup.*vs*.Ctrl), 287 WBS DEGs (Del.*vs*.Ctrl), with 153 in common and

a fold change expression over controls in the same direction (Figure 3A). When we compared DEGs from Dup7 against WBS fibroblast lines (Dup.*vs.*Del), we identified 28 significant DEGs. Enrichment analysis revealed specific pathways and GO categories relevant for the hallmark phenotypes of both disorders (Figure 3B). In Dup7, the only pathway reaching statistical significance after FDR correction was Arf6 trafficking events (q-value=0.0313). ARF6 is a small GTPase that participates in intracellular vesicular trafficking, endocytosis processes and actin remodelling (51). The DEGs from the WBS fibroblast lines were enriched in cellular pathways related to PI3K-Akt signaling, extracellular matrix organization and cell adhesion. Additionally, GO categories for both Dup7 and WBS lines included categories involving the development of important organs that are affected in the two conditions such as cardiovascular system, connective tissue, nervous system, skeletal system and kidney development (Figure 3B).

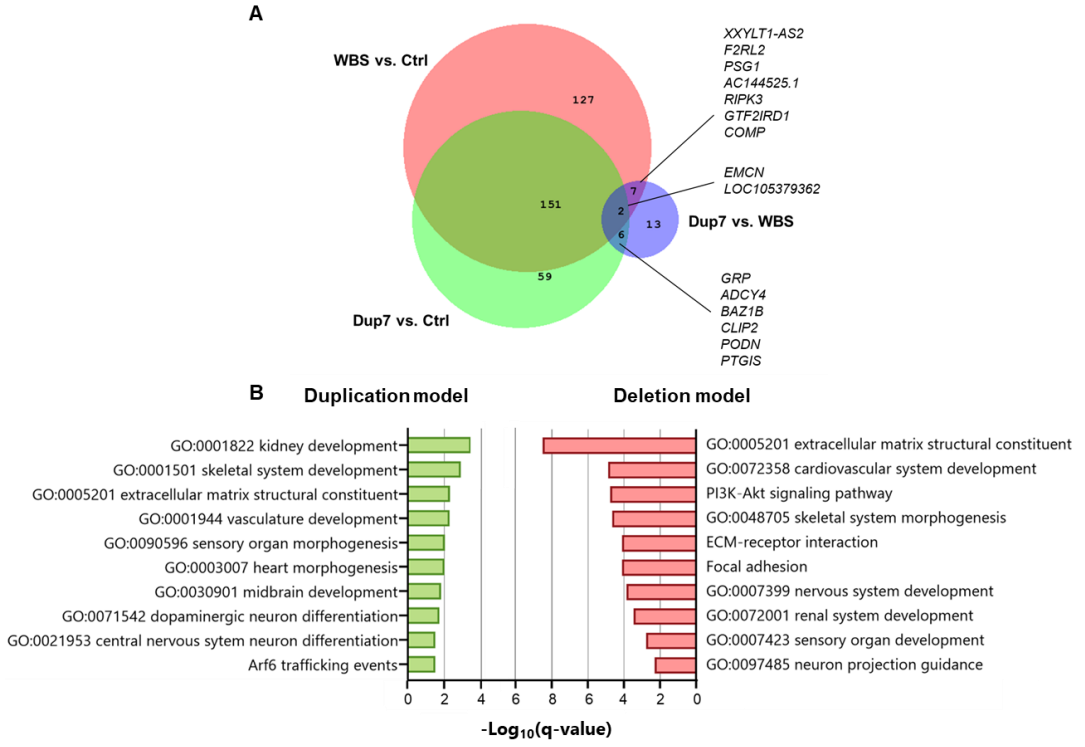


Figure 3. Differential gene expression and enrichment analysis in fibroblasts. (A) Number and distribution of DEGs across the three 7q11.23 groups. (B) Selection of GO categories relevant for hallmark phenotypes of the two disorders and significant after FDR correction for multiple testing.

iPSCs and NPCs

The number of DEGs identified in iPSCs was smaller. We detected 7 DEGs in the duplication model (Dup.*vs*.Ctrl), 20 DEGs in the deletion model (Del.*vs*.Ctrl) and 51 DEGs in the comparison between Dup7 and WBS (Dup.*vs*.Del) (Supplementary Figure 6A). Enrichment analysis identified an overrepresentation of the Gene expression (Transcription) pathway (q-value=0.0134) in Dup7 DEGs, and GO categories related to transcription, RNA metabolism and regulation of gene expression in WBS DEGs.

In NPCs, the detected DEGs were limited; 3 DEGs in WBS lines and 42 DEGs in Dup.*vs*.Del categories, whereas no DEGs were found in Dup7 NPC lines compared to controls (Supplementary Figure 6B). Among the DEGs resulting from the duplicated and deleted models comparison, there was one pathway commonly overrepresented in both iPSCs and NPCs, the Basal transcription factors pathway, led by the 7q11.23 genes *GTF2I* and *GTF2IRD1*.

Neurons

In iNeu, we detected 36 DEGs in Dup7 lines (Dup.*vs*.Ctrl), 7 DEGs in WBS (Del.*vs*.Ctrl) and 23 DEGs when comparing duplicated against deleted models (Dup.*vs*.Del). Eleven genes were shared between two of these comparisons (Figure 4A). Interestingly, seven of them are highly expressed in brain according to GTEx Portal and The Human Protein Atlas, and four have relevant functions in neurons (52). In addition, some have been previously linked to ASD including *CNTNAP4* (contactin associated protein family member 4) and *NXPH1* (neurexophilin 1). Remarkably, pathway and GO analysis of Dup7 DEGs uncovered significant enrichments for neuronal processes and pathways, and the top-ranking signaling pathways included transmission across chemical synapses, neuronal system and GABAergic synapse (Figure 4B). However, these results should be interpreted with caution taking into account the limited sample size of this cell type.

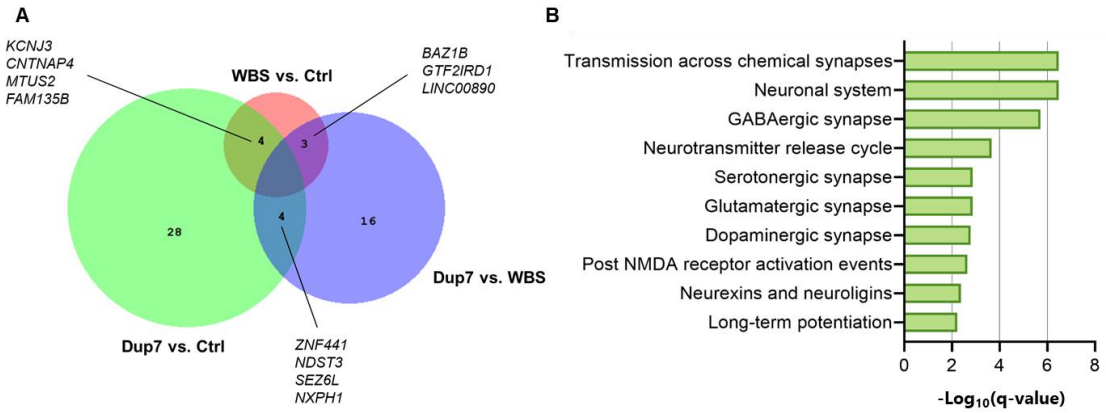


Figure 4. Differential gene expression and enrichment analysis in iNeus. (A) Number and distribution of DEGs across the three 7q11.23 genotype groups in neurons. (B) Brain-specific pathways enriched among DEGs in Dup7 vs. Ctrl and significant after FDR correction.

WGCNA reveals co-expression modules correlated with 7q11.23 copies

In order to explore the transcriptional changes in a system-level context and to identify co-expression modules associated with the duplication and deletion models, we performed a WGCNA approach in iPSCs and NPCs cell types, separately, as they were the cell lines with enough sample size. In iPSCs, our analysis detected 27 co-expression modules (Figure 5). Three of these modules, $M1_{iPSC}$ (brown4), $M15_{iPSC}$ (maroon) and $M25_{iPSC}$ (skyblue3) were significantly correlated with the genotypes derived from imbalances in 7q11.23 dosage. $M1_{iPSC}$ (r=0.48, p-value=0.04) contained mainly upregulated genes in Dup7, while WBS genes were downregulated. This module showed enrichment for genes related to cholesterol biosynthesis as well as transmembrane and ion transport. In addition, one module-hub gene (*SLC7A5*) was reported in SFARI database. $M15_{iPSC}$ and $M25_{iPSC}$ presented a significant inverse correlation (r=-0.54, p-value=0.02 and r=-0.47, p-value=0.04, respectively), with genes mainly downregulated in Dup7 and upregulated in WBS. The majority of the module-hub genes in $M15_{iPSC}$ were classified as small nucleolar RNAs and $M25_{iPSC}$ was enriched in mesenchymal cell development as well as other GO categories related to DNA binding and histone deacetylase activity.

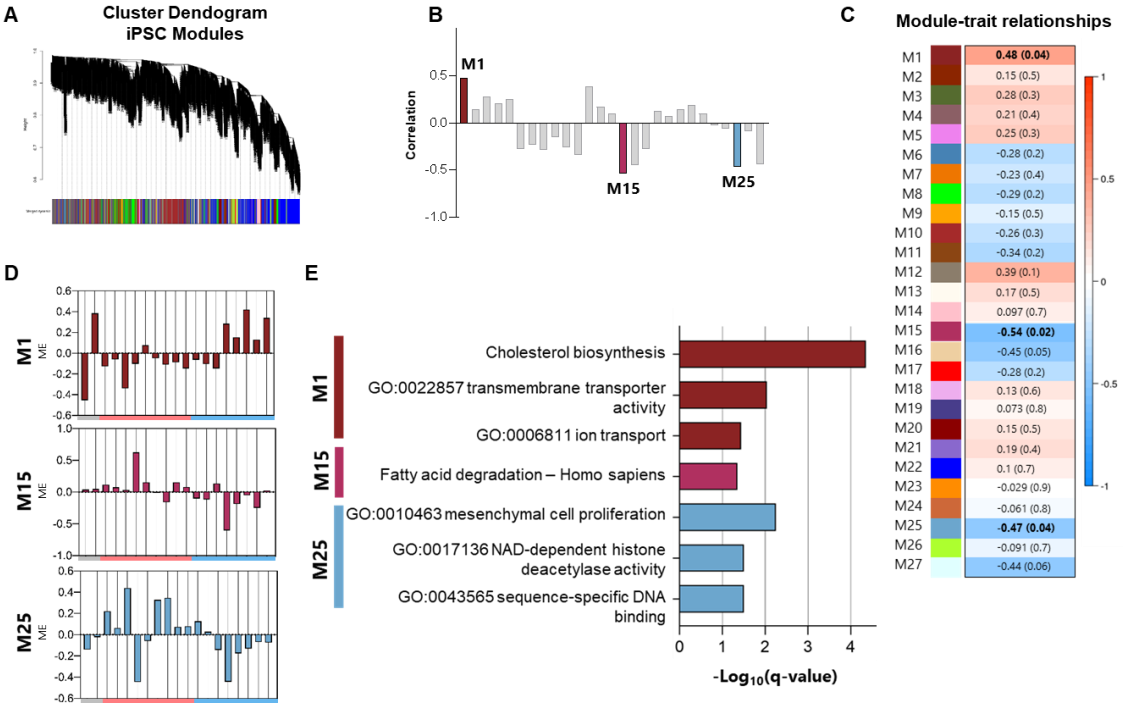


Figure 5. Weighted gene co-expression network analysis (WGCNA) for iPSCs. (A) Dendogram showing the 27 co-expression modules detected in iPSCs. (B) Correlation analysis between the module eigengenes (MEs) and 7q11.23 genotype. The coloured bars are modules associated with the trait. (C) A heatmap of the positive (red) and negative (blue) associations between modules and 7q11.23 copy number. (D) Bar plots of ME expression for all significantly associated modules. The color below the plot indicates the sample group: grey for control samples, red for WBS patients and blue for Dup7. (E) Three representative pathways and/or GO terms for each module associated with the 7q11.23 genotype (q-value<0.05).

In the WGCNA study of the NPC lines, we identified 42 co-expression modules, from which five were significantly correlated with 7q11.23 copies (Figure 6). From these, only $M25_{NPC}$ (green) was positively associated with the copy number genotype (r=0.5, p-value=0.03) and the genes included appeared mostly upregulated in Dup7 and downregulated in WBS. This module was enriched with genes related to connexin binding. The other four modules presented a significant inverse correlation, which included genes largely downregulated in Dup7 and upregulated in WBS. The highest negative correlation was observed in $M14_{NPC}$ (ivory, r=-0.77, p-value=1e-04), without significant results in the enrichment analysis. Notably, four module-hub genes were located within the 7q11.23 duplicated/deleted region (*CLIP2, BAZ1B, BCL7B* and *DNAJC30*). $M4_{NPC}$ (honeydew1, r=-0.54, p-value=0.02) was composed of genes involved in RNA

binding and ribonucleoprotein complex assembly. Of note is the identification of two ASD strong candidates reported in SFARI (*PRKD1* and *SATB1*) among the module-hub genes in $M8_{NPC}$ module (firebrick3, r=-0.73, p-value=4e-04). Remarkably, $M41_{NPC}$ module (orangered3, r=-0.5, p-value=0.03) was enriched in genes related to regulation of signaling pathways, including glutamate receptor signaling. Glutamatergic system is implicated in basic functions in the central nervous system, and its dysfunction has been recurrently associated with ASD (53).

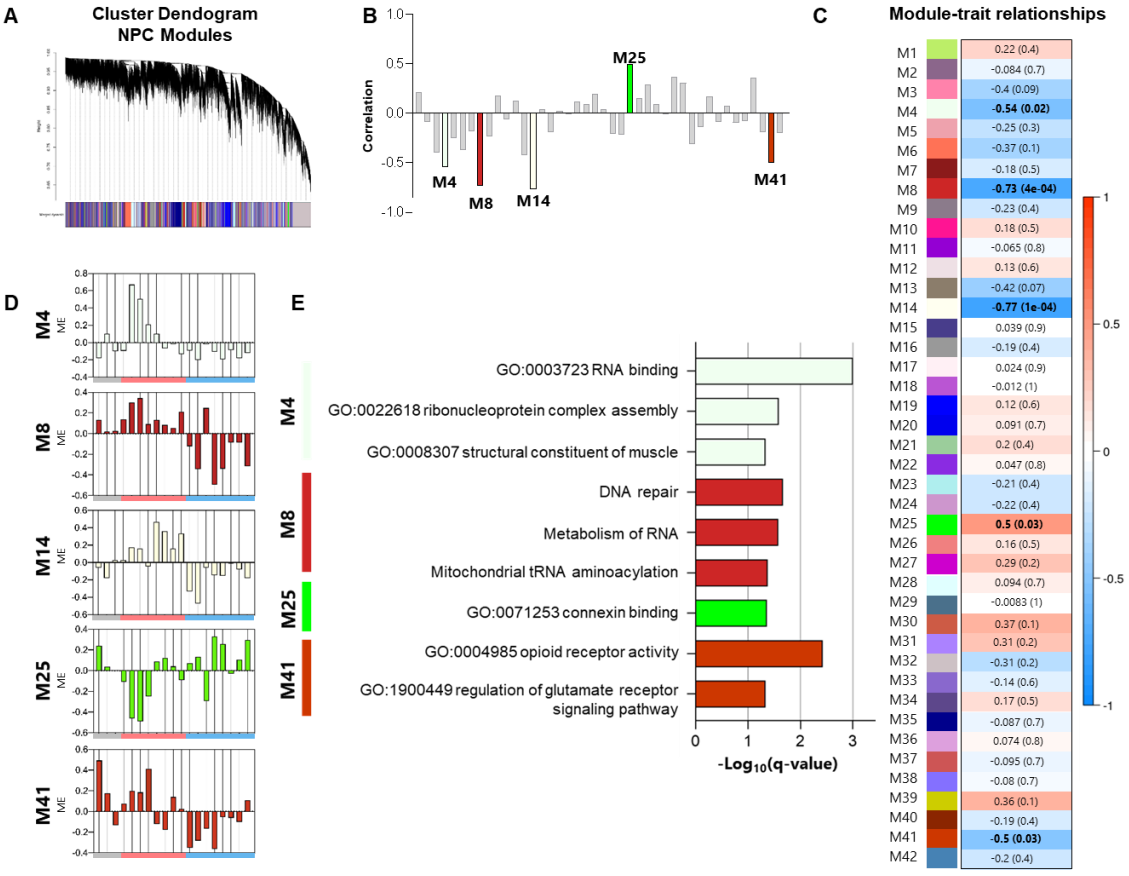


Figure 6. Weighted gene co-expression network analysis (WGCNA) analyses for NPC lines. (A) Dendogram showing the 42 co-expression modules detected in NPCs. (B) Correlation analysis between the module eigengenes (MEs) and 7q11.23 genotype. The bars are modules associated with the trait. (C) A heatmap of the positive (red) and negative (blue) associations between modules and 7q11.23 copy number. (D) Bar plots of ME expression for all significantly associated modules. The color below the plot indicates the sample group: grey for control samples, red for WBS patients and blue for Dup7. (E) The top enriched pathways and/or GO terms for each module associated with the 7q11.23 genotype (q-value<0.05).

DISCUSSION

We present here one of the largest 7q11.23 rearrangements patient-specific iPSCs cohorts, including cellular lines derived from four Dup7 and four WBS patients. Considering that both conditions are neurodevelopmental disorders and given the difficulty to study the developing brain, iPSC-based technologies are a good approach to gain insight into the transcriptomic consequences of 7q11.23 imbalances in multiple *in vitro* differentiation stages (25).

The majority of the patients in our cohort were carriers of 7q11.23 microduplication or microdeletion without additional copy number alteration at other rare CNV *loci* (50). However, patient WBS159 and its derivative models, were also carrying a *de novo* 16p13.11 microduplication. The 16p13.11 microduplication is a rare but recurrent CNV *loci* with a prevalence in the general population of 0.13-0.17% (50,54). Although it is a variant with low penetrance (8.4%), this CNV is frequently associated with ID, developmental delay, ASD and schizophrenia (50,54,55). Neuronal induction for this line was particularly problematic and required multiple repeated attempts. Similarly, a recent study found a significant reduction of NPC proliferation in 16p13.11 microduplication patient iPSC-derived cells compared to controls (56). Thus, the presence of a second hit CNV in WBS159 could be the cause of the additional troubled steps in NPC induction. Beside this particularly difficult case, we were able to successfully generate and characterize NPCs from iPSCs, and iNeu (for a subset of samples) to assess the genome-wide transcriptional alterations.

The expression of the genes in the 7q11.23 single-copy region showed two tendencies: a symmetrically opposite expression and an alteration in the same direction. These findings were consistent with previous transcriptional studies of 7q11.23 iPSC-based cells in 92-96% of the comparisons in fibroblasts (22), iPSCs (26,28) and/or derived neurons (27,28). From the 30 genes assessed in the 7q11.23 interval, 11 showed a uniform upregulation in Dup7 (iPSCs) and downregulation in WBS (fibroblasts, iPSCs and neurons). These concordances

confirm the regular gene expression pattern of the 7q11.23 genes, previously described in blood transcriptome samples from Dup7 and WBS patients (21,24,57,58). Specifically, *BAZ1B, BCL7B, TBL2, WBSCR22, EIF4H, RFC2, CLIP2* and GTF2IRD1 have been consistently deregulated in all studies and cell types, suggesting they are more dose sensitive and/or that fewer *trans* feedback mechanisms are acting to normalize their expression.

Among the 7q11.23 DEGs, we highlighted the potential player *MIR590* that presented the highest differential expression between Dup.vs.Del in iPSCs, NPCs and iNeus. This differentially expressed miRNA was previously associated with cardiomyocyte proliferation and anti-apoptotic functions in atherosclerosis (59,60), but it has also been described as a tumor suppressor gene in a variety of cancers (61,62). We have found that *MIR590* predicted mRNA targets are enriched in ASD risk genes from SFARI and in genes altered in WBS patients by transcriptomic analysis, suggesting that this miRNA could regulate the expression of genes important for both disorders. Interestingly, some potential target genes of hsa-miR-590-5p/3p were found deregulated in iNeu, including *SLC5A7, SPOCK3, NDST3* and *BCL7B*. In particular, *SLC5A7* (solute carrier family 5 member 7) is involved in neurotransmission, a process consistently found to be altered in Dup7 and WBS (2,27). In fact, miRNAs are gaining more attention in the field of neurodevelopmental disorders, especially in ASD. Many studies have identified significantly deregulated miRNA expression in multiple tissues from ASD patients, although *MIR590* has not been reported before (63). Some of the recurrently altered miRNAs in ASD include miR-23a, miR-132 and miR-146b (64–66) and their predicted targets were also enriched for ASD susceptibility genes (67). We believe that experimental confirmation of the hsa-miR-590-5p/3p targets warrants further investigation and would be valuable to determine whether *MIR590* could be a therapeutic target for 7q11.23 CNV patients.

Assessing genome-wide differential expression, we discovered 494 unique DEGs. For the Dup.vs.Del comparisons, we intersected the lists of DEGs in the different cell types, and we observed three genes outside the 7q11.23 *locus* that

were differentially expressed in multiple cell types: *ZNF726, ZNF439* and *LOC105374122*. Both *ZNF726* and *ZNF439* are zinc finger proteins that have been detected to be methylated in cancer samples and are probably involved in transcriptional regulation (68,69). *ZNF726* and *ZNF439* may be relevant to the phenotype in early development and these findings suggest considering the study of epigenetic changes in Dup7 and WBS in the future.

Transcriptional consequences in fibroblasts affected a great proportion of DEGs. Arf6 trafficking events was the only pathway significantly enriched among Dup7 DEGs and it is involved in actin cytoskeleton remodelling and endosomal membrane trafficking with important functions in neuronal development and synaptic plasticity (51,70). Arf6 is a small GTP-binding protein expressed in the nervous system and concretely in the brain, where evidence showed its participation in regulating dendritic formation, axonal growth and synaptic plasticity (70). Besides, a missense mutation in *IQSEC2* (IQ motif and Sec7 domain ArfGEF2) gene was reported in a patient with ID, autism and epilepsy. They demonstrated *in vitro* that this mutation resulted in a constitutive activation of Arf6 that could lead to a reduction in surface AMPA receptors providing a mechanistic link between the mutation and the behavioral and cognitive phenotype of the patient (71). Taken together, it suggests that impairments of this pathway could partially contribute to the autistic phenotype of Dup7 individuals.

From the 287 DEGs detected in WBS model (Del.vs.Ctrl), 27/868 (3.1%) were also found deregulated in Henrichsen *et al.* (2011) (22), including 77.7% (21/27) in the same direction. The small overlap between studies is probably due to limited sample sizes (8 WBS patients in Henrichsen *et al*. and 4 in this study), reinforcing the relevance of this set of overlapping genes and the need of replication studies. Among these genes we observe *COL4A2* (collagen type IV alpha 2 chain) and *EFNA5* (ephrin A5) that are involved in PI3k-Akt signaling pathway, which has been proposed to be compromised in WBS (72). GO analysis revealed categories related to the hallmark phenotypes of both disorders, and

similar results were reported by Adamo *et al.* (26) in iPSC lines from Dup7 and WBS individuals.

Differentially expressed genes in iPSCs were related to regulation of transcription and gene expression, partially due to 7q11.23 genes (*GTF2IRD1* and *BAZ1B*), but also driven by the alteration of other transcriptional regulators such as *ZNF71, ZNF454, ZNF676* and *BAZ1A*. Interestingly, *BAZ1A* belongs to a bromodomain family together with *BAZ1B, BAZ2A* and *BAZ2B*, and encodes the accessory subunit of ACF1 (ATP-dependent chromatin assembly factor), which is a member of the ISWI ("imitation switch") chromatin-remodeling complexes (73). *De novo* missense mutations and microdeletions in this gene have been associated with syndromic ID and brain malformations, and therefore it was proposed to be involved in neurodevelopment (74–76).

Our systems biology approach in iPSCs revealed three co-expression modules significantly correlated with 7q11.23 dosage, and specifically, $M1_{iPSC}$ module, containing mainly genes upregulated in Dup7 and downregulated in WBS, was enriched in genes related to cholesterol biosynthesis and ion transport. Evidence suggests an alteration of cholesterol homeostasis among some ASD patients (77–79). Total cholesterol levels were higher in Asperger syndrome individuals compared to controls (77), whereas patients with Smith-Lemli-Opitz syndrome (SLOS) showed low cholesterol plasma levels (79). SLOS is caused by mutations in *DHCR7* (7-dehydrocholesterol reductase), a gene involved in cholesterol biosynthesis, and it is considered a syndromic form of ASD as in 53-71% of cases comorbid ASD is reported (80–82). It has been hypothesized that lipid raft disarrangements, in which cholesterol is a key component, could alter synaptic function, providing a link between cholesterol metabolism and ASD (78). In contrast, low-normocholesterolemia has been observed among WBS patients (83), suggesting that alteration of cholesterol biosynthesis could have more relevance in Dup7 individuals than in WBS subjects. Finding alterations in the pluripotent state indicates that cells from 7q11.23 CNVs patients carry changes from a very early developmental stage, which can have greater consequences in the differentiated lineages.

The WGCNA analysis in NPCs uncovered five co-expression modules significantly associated with 7q11.23 copy number genotype. We point out $M8_{NPC}$ and $M41_{NPC}$, where the number of 7q11.23 copies was inversely related to expression intensity. Two module-hub genes in $M8_{NPC}$ have been strongly linked to ASD: *PRKD1* (protein kinase D1) and *SATB1* (SATB homeobox 1) (84,85). Decreased expression of *PRKD1* was observed in three patients with atypical Rett syndrome bearing a *de novo* 14q12 deletion that includes this gene (84), whereas *SATB1* was identified as an ASD candidate gene in a recent TADA (transmitted and *de novo* analysis) study (85). In the $M41_{NPC}$ module, the regulation of glutamate receptor signaling pathway was enriched among its components. There is increasing evidence supporting a glutamate dysfunction in the pathophysiology of ASD (53,86–90). First, glutamate levels have been reported to be significantly higher in ASD patients compared to controls both in blood and brain (53). In addition, microarray analyses have observed alterations in different members of the glutamate system (86). For instance, increased mRNA levels of AMPA-type glutamate receptors and glutamate transporters were identified in post-mortem brain tissues of ASD patients (87). Lastly, not only many ASD risk genes are related to glutamatergic synapses, such as *GRIN2B, GRIA2* or *SLC1A2,* but also genes in ASD-specific CNVs are enriched in glutamatergic synapse in different studies (53,88–90). Therefore, the glutamate receptor signaling could be important for the autistic phenotype of Dup7 and WBS patients, since ASD features present a 6-10-fold increase in WBS subjects compared to general population.

In iNeus, we detected eleven genes that were present in different DEG comparisons (Figure 4A), of which a considerable proportion were enhanced in brain and with important functions in the nervous system. Specifically, *CNTNAP4* and *NXPH1* are ASD risk genes reported in SFARI database. *CNTNAP4* is a presynaptic protein from the neurexin family involved in dopaminergic and GABAergic synaptic transmission that was downregulated in both Dup7 and WBS models. Deletions disrupting this gene have been identified in ASD patients in several studies (91,92). *NXPH1* acts binding α-neurexins to promote adhesion between dendrites and axons (93). In this case,

familial CNVs (both deletions and duplications) disrupting *NXPH1* have been found in ASD individuals (94,95). It is also worth noting two additional genes: *KCNJ3* (potassium inwardly rectifying channel subfamily J member 3) and *SEZ6L* (seizure related 6 homolog like). In our study, *KCNJ3* was downregulated in both Dup7 and WBS lines compared to controls, and the same observation was detected in Fragile X syndrome iPSC-derived neurons, another syndromic form of ASD (96). A missense mutation was identified in *SEZ6L* in an exome sequencing study of ASD patients and it was considered as an exceptionally strong ASD candidate in a recent association test (97). Although these results should be cautiously interpreted, enriched pathways in DEGs from Dup7 model might indicate an alteration of synaptic function in Dup7 neurons probably affecting the establishment of cerebral patterns since very early in the neurodevelopment. Although in our study, WBS iNeus did not show a high number of DEGs or clear altered pathways, abnormalities in neuronal gene expression have been described in WBS derived-neurons (27).

Limitations

Our study presents several limitations. First, in terms of sample selection, the number of controls in our cohort is half the number of cases' samples, which restricted our statistical power in the comparisons against controls. Second, the results in iNeus should be interpreted with caution because only half of the samples were differentiated from NPCs into iNeus due to economic constraints. Third, our experiments were performed in two experimental rounds, which could increase the transcriptional differences among samples, but no batch effect was detected. Fourth, although iPSCs are a good approach to study developmental brain disorders, these *ex-vivo* models do not precisely reproduce all the events occurring during brain development. Fifth, the comparison with previous iPSCs-based transcriptomic studies should be interpreted cautiously due to experimental differences, including iPSCs generation methods and a diversity of differentiation protocols. Finally, a general limitation of iPSCs studies is the reduced sample size due to high cost in economic and temporal terms. Considering all 7q11.23 iPSC-derived studies we count: 25 WBS iPSCs lines, 2 atypical WBS iPSCs lines, and 7 Dup7 iPSC lines

(22,26–30). This is why, it would be truly interesting to develop and conduct a meta-analysis for iPSCs (and derivatives) in a collaborative effort in the scientific community.

Conclusions

To sum up, we present one of the largest 7q11.23 patient-derived iPSC lines and derivatives including both duplication and deletion models. Gene expression from 7q11.23 region has proved to be consistent, further supporting expression imbalances of *BAZ1B, BCL7B, TBL2, WBSCR22, EIF4H, RFC2, CLIP2* and GTF2IRD1, and highlighting *MIR590* as a novel compelling candidate to regulate the expression of genes relevant for both Dup7 and WBS disorders. Integrative transcriptomic analysis of *in vitro* 7q11.23-CNVs cellular models revealed genes and pathways related to cholesterol metabolism, Arf6 trafficking events and glutamate signaling altered during early neuronal development in these genomic disorders, which could lead to novel potential therapeutic targets.

Funding

This work was funded by grants from Fundación Ramon Areces (CIVP16A1828), the marathon "Todos Somos Raros, Todos Somos Únicos" (P52), the Spanish Ministry of Science, Innovation and Universities (RTI2018-101960-A-I00) and the Spanish Ministry of Economy and Competiveness (FIS16/00369). RC was recipient of a Marie Skłodowska-Curie Action H2020 (656369) and Ramón y Cajal (RYC-2017-21636) fellowships. MCR had a predoctoral fellowship of the Ministry of Education, Culture and Sport (FPU16/06907). Also, Red de Terapia Celular, TerCel, funded by Retics, **RD12/0019/0034**; Plataforma de Recursos Biomoleculares y Bioinformaticos, **PRB2, PT13/0001/0041**; both funded by Instituto de Salud Carlos III-ISCIII/FEDER.

Authors' contributions

MC-R performed all the transcriptomic analysis. BK, RF, NU, AA-M, MA-E, AV, IC and RC participated in the experiments. MC-R, IC, RC and LAPJ drafted the manuscript. IC, RC and LAPJ conceived the study and participated in the design.

REFERENCES

1. Carvalho CMB, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nat Revs Genet. 2016;17:224–38.
2. Merla G, Brunetti-Pierri N, Micale L, Fusco C. Copy number variants at Williams-Beuren syndrome 7q11.23 region. Hum Genet. 2010;128:3–26.
3. 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;69(1):1–13.
4. 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;66(1):47–68.
5. Bayés M, Magano LF, Rivera N, Flores R, Pérez Jurado LA. Mutational mechanisms of Williams-Beuren syndrome Deletions. Am J Hum Genet. 2003;73(1):131–51.
6. Strømme P, Bjømstad PG, Ramstad K. Prevalence estimation of Williams syndrome. J Child Neurol. 2002;17(4):269–71.
7. Pober B. Williams-Beuren Syndrome. N Engl J Med. 2010;362:239–52.
8. Mervis CB, Klein-tasman BP. Williams Syndrome: Cognition, personality and adaptive behavior. MRDD Res Rev. 2000;158(6):148–58.
9. 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. 2009;52:94–100.
10. 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;18(3):108–16.
11. 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;167A(12):2916–35.
12. Abbas E, Cox D, Smith T, Butler M. The 7q11.23 Microduplication Syndrome: A Clinical Report with Review of Literature. J Pediatr Genet. 2016;5:129–40.
13. Klein-Tasman BP, van der Fluit F, Mervis CB. Autism Spectrum Symptomatology in Children with Williams Syndrome Who Have Phrase Speech or Fluent Language. J Autism Dev Disord. 2018;48(9):3037–50.
14. 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;75(5):371–7.
15. 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;70(5):863–85.
16. Ewart AK, Morris CA, Atkinson D, Jin W, Sternes K, Spallone P, et al. Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. Nat Genet.1993;5(1):11–6.
17. Morris CA, 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.
18. Serrano-Juárez CA, Prieto-Corona B, Rodríguez-Camacho M, Venegas-Vega CA, Yáñez-Téllez MG, Silva-Pereyra J, et al. An Exploration of Social Cognition in Children with Different Degrees of Genetic Deletion in Williams Syndrome. J Autism Dev Disord 2020.
19. Lugo M, Wong ZC, Billington CJ, Parrish PCR, Muldoon G, Liu D, et al. Social, neurodevelopmental, endocrine, and head size differences associated with atypical deletions in Williams–Beuren syndrome. Am J Med Genet Part A. 2020;182(5):1008–20.
20. Delio M, Pope K, Wang T, Samanich J, Haldeman-Englert CR, Kaplan P, et al. Spectrum

of elastin sequence variants and cardiovascular phenotypes in 49 patients with Williams-Beuren syndrome. Am J Med Genet Part A. 2013;161(3):527–33.
21. Antonell A, Vilardell M, Jurado LAP. Transcriptome profile in Williams-Beuren syndrome lymphoblast cells reveals gene pathways implicated in glucose intolerance and visuospatial construction deficits. Hum Genet. 2010;128(1):27–37.
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. Shamir R, editor. PLoS Comput Biol. 2011;7(1):e1001054.
23. Benítez-Burraco A, Kimura R. Robust Candidates for Language Development and Evolution Are Significantly Dysregulated in the Blood of People With Williams Syndrome. Front Neurosci. 2019;13:1–9.
24. Kimura R, Swarup V, Tomiwa K, Gandal MJ, Parikshak NN, Funabiki Y, et al. Integrative network analysis reveals biological pathways associated with Williams syndrome. J Child Psychol Psychiatry Allied Discip. 2019;60(5):585–98.
25. Iakoucheva LM, Muotri AR, Sebat J. Getting to the Cores of Autism. Cell. 2019;178(6):1287–98.
26. 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;47(2):132–41.
27. Khattak S, Brimble E, Zhang W, Zaslavsky K, Strong E, Ross PJ, et al. Human induced pluripotent stem cell derived neurons as a model for Williams-Beuren syndrome. Mol Brain. 2015;8(1):1–11.
28. Lalli MA, Jang J, Park JHC, Wang Y, Guzman E, Zhou H, et al. Haploinsufficiency of BAZ1B contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways. Hum Mol Genet. 2016;25(7):1294–306.
29. Chailangkarn T, Trujillo CA, Freitas BC, Hrvoj-mihic B, Herai RH, Yu DX, et al. A human neurodevelopmental model for Williams syndrome. Nature. 2016;536(7616):338–43.
30. Zanella M, Vitriolo A, Andirko A, Martins PT, Sturm S, O'Rourke T, et al. Dosage analysis of the 7q11.23 Williams region identifies BAZ1B as a major human gene patterning the modern human face and underlying self-domestication. Sci Adv. 2019;5(12):eaaw7908.
31. Kuebler B, Aran B, Flores R, Pérez-Jurado LA, Veiga A, Cuscó I, et al. Derivation of induced pluripotent stem cells (iPSCs) by retroviral transduction of skin fibroblasts from four patients suffering 7q11.23 microduplication syndrome. Stem Cell Res. 2020;49:102092.
32. Kuebler B, Aran B, Flores R, Pérez-Jurado LA, Veiga A, Corominas R, et al. Generation of induced pluripotent stem cells (iPSCs) by retroviral transduction of skin fibroblasts from four patients suffering Williams-Beuren syndrome (7q11.23 deletion). Stem Cell Res. 2020;49:102087.
33. Yan Y, Shin S, Jha BS, Liu Q, Sheng J, Li F, et al. Efficient and Rapid Derivation of Primitive Neural Stem Cells and Generation of Brain Subtype Neurons From Human Pluripotent Stem Cells. Stem Cells Transl Med. 2013;2(11):862–70.
34. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):R80.
35. The Comprehensive R Archive Network https://cran.r-project.org/.
36. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4(2):249–64.
37. Bengtsson H, Irizarry R, Carvalho B, Speed TP. Estimation and assessment of raw copy numbers at the single locus level. Bioinformatics. 2008;24(6):759–67.
38. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
39. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC

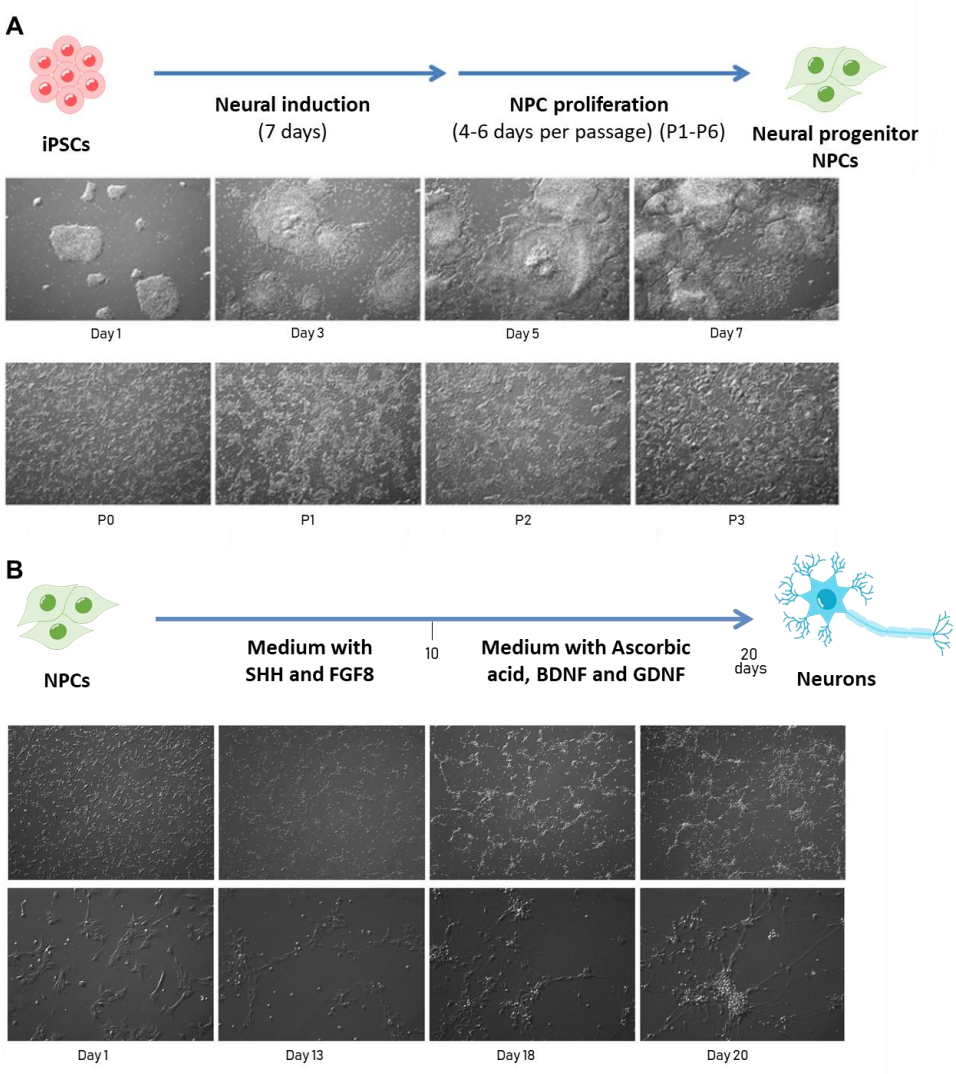
Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34:D590-8.
40. Hulsen T, de Vlieg J, Alkema W. BioVenn - A web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. BMC Genomics. 2008;9:488.
41. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB - A database for integrating human functional interaction networks. Nucleic Acids Res. 2009;37:D623-8.
42. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523.
43. 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.
44. Codina-Sola M, Costa-Roger M, Pérez-García D, Flores R, Palacios-Verdú MG, Cusco I, et al. Genetic factors contributing to autism spectrum disorder in Williams-Beuren syndrome. J Med Genet. 2019;0:1–8.
45. Vejnar CE, Zdobnov EM. MiRmap: Comprehensive prediction of microRNA target repression strength. Nucleic Acids Res. 2012;40(22):11673–83.
46. Sticht C, De La Torre C, Parveen A, Gretz N. Mirwalk: An online resource for prediction of microrna binding sites. PLoS One. 2018;13(10):e0206239.
47. Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, et al. MiRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. Nucleic Acids Res. 2018;46(D1):D296–302.
48. Lu TP, Lee CY, Tsai MH, Chiu YC, Hsiao CK, Lai LC, et al. MiRSystem: An integrated system for characterizing enriched functions and pathways of microRNA targets. PLoS One. 2012;7(8):e42390.
49. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
50. Malhotra D, Sebat J. CNVs: Harbingers of a rare variant revolution in psychiatric genetics. Cell. 2012;148(6):1223–41.
51. D'Souza-Schorey C, Chavrier P. ARF proteins: Roles in membrane traffic and beyond. Nat Rev Mol Cell Biol. 2006;7(5):347–58.
52. Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, et al. The GeneCards suite: From gene data mining to disease genome sequence analyses. Curr Protoc Bioinforma. 2016;54:1.30.1-1.30.33.
53. Rojas DC. The role of glutamate and its receptors in autism and the use of glutamate receptor antagonists in treatment. J Neural Transm. 2014;121(8):891–905.
54. El Khattabi LA, Heide S, Caberg JH, Andrieux J, Doco Fenzy M, Vincent-Delorme C, et al. 16p13.11 microduplication in 45 new patients: Refined clinical significance and genotype-phenotype correlations. J Med Genet. 2020;57(5):301–7.
55. Ingason A, Rujescu D, Cichon S, Sigurdsson E, Sigmundsson T, Pietiläinen OPH, et al. Copy number variations of chromosome 16p13.1 region associated with schizophrenia. Mol Psychiatry. 2011;16(1):17–25.
56. Johnstone M, Vasistha NA, Barbu MC, Dando O, Burr K, Christopher E, et al. Reversal of proliferation deficits caused by chromosome 16p13.11 microduplication through targeting NFκB signaling: an integrated study of patient-derived neuronal precursor cells, cerebral organoids and in vivo brain imaging. Mol Psychiatry. 2019;24(2):294–311.
57. Merla G, Howald C, Henrichsen CN, Lyle R, Wyss C, Zabot MT, et al. Submicroscopic deletion in patients with Williams-Beuren syndrome influences expression levels of the nonhemizygous flanking genes. Am J Hum Genet. 2006;79(2):332–41.
58. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, et al. Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am J Hum Genet. 2012;91(1):38–55.
59. Eulalio A, Mano M, Ferro MD, Zentilin L, Sinagra G, Zacchigna S, et al. Functional scre-

ening identifies miRNAs inducing cardiac regeneration. Nature. 2012;492(7429):376–81.
60. Bao MH, Li JM, Zhou QL, Li GY, Zeng J, Zhao J, et al. Effects of miR-590 on oxLDL-induced endothelial cell apoptosis: Roles of p53 and NF-κB. Mol Med Rep. 2016;13(1):867–73.
61. Wang WT, Qi Q, Zhao P, Li CY, Yin XY, Yan RB. miR-590-3p is a novel microRNA which suppresses osteosarcoma progression by targeting SOX9. Biomed Pharmacother. 2018;107:1763–9.
62. Xu BB, Gu ZF, Ma M, Wang JY, Wang HN. MicroRNA-590-5p suppresses the proliferation and invasion of non-small cell lung cancer by regulating GAB1. Eur Rev Med Pharmacol Sci. 2018;22(18):5954–63.
63. Wu X, Li W, Zheng Y. Recent Progress on Relevant microRNAs in Autism Spectrum Disorders. Int J Mol Sci. 2020;21:5904.
64. Abu-Elneel K, Liu T, Gazzaniga FS, Nishimura Y, Wall DP, Geschwind DH, et al. Heterogeneous dysregulation of microRNAs across the autism spectrum. Neurogenetics. 2008;9(3):153–61.
65. Sarachana T, Zhou R, Chen G, Manji HK, Hu VW. Investigation of post-transcriptional gene regulatory networks associated with autism spectrum disorders by microRNA expression profiling of lymphoblastoid cell lines. Genome Med. 2010;2(4):23.
66. Talebizadeh Z, Butler MG, Theodoro MF. Feasibility and relevance of examining lymphoblastoid cell lines to study role of microRNAs in autism. Autism Res. 2008;1(4):240–50.
67. Constantin L. The Role of MicroRNAs in Cerebellar Development and Autism Spectrum Disorder During Embryogenesis. Mol Neurobiol. 2017;54(9):6944–59.
68. Zhang H, Sun X, Lu Y, Wu J, Feng J. DNA-methylated gene markers for colorectal cancer in TCGA database. Exp Ther Med. 2020;19(4):3042–50.
69. Ren S, Gaykalova D, Wang J, Guo T, Danilova L, Favorov A, et al. Discovery and development of differentially methylated regions in human papillomavirus-related oropharyngeal squamous cell carcinoma. Int J Cancer. 2018;143(10):2425–36.
70. Jaworski J. ARF6 in the nervous system. Eur J Cell Biol. 2007 Sep 14;86(9):513–24.
71. Rogers EJ, Jada R, Scharagenheim-Rozales K, Sah M, Cortes M, Florence M, et al. An IQSEC2 Mutation Associated With Intellectual Disability and Autism Results in Decreased Surface AMPA Receptors. Front Mol Neurosci. 2019;12(43):1–18.
72. Segura-Puimedon M, Borralleras C, Pérez-Jurado LA, Campuzano V. TFII-I regulates target genes in the PI-3K and TGF-β signaling pathways through a novel DNA binding motif. Gene. 2013;527(2):529–36.
73. Jones MH, Hamana N, Nezu JI, Shimane M. A novel family of bromodomain genes. Genomics. 2000;63(1):40–5.
74. Zaghlool A, Halvardson J, Zhao JJ, Etemadikhah M, Kalushkova A, Konska K, et al. A role for the chromatin-remodeling factor BAZ1A in neurodevelopment. Hum Mutat. 2016;37(9):964–75.
75. Weitensteiner V, Zhang R, Bungenberg J, Marks M, Gehlen J, Ralser DJ, et al. Exome sequencing in syndromic brain malformations identifies novel mutations in ACTB, and SLC9A6, and suggests BAZ1A as a new candidate gene. Birth Defects Res. 2018;110(7):587–97.
76. Ponzi E, Gentile M, Agolini E, Matera E, Palumbi R, Buonadonna A, et al. 14q12q13.2 microdeletion syndrome: Clinical characterization of a new patient, review of the literature, and further evidence of a candidate region for CNS anomalies. Mol Genet Genom Med. 2020;8:e1289.
77. Dziobek I, Gold SM, Wolf OT, Convit A. Hypercholesterolemia in Asperger syndrome: Independence from lifestyle, obsessive-compulsive behavior, and social anxiety. Psychiatry Res. 2007;149(1–3):321–4.
78. Wang H. Lipid rafts: a signaling platform linking cholesterol metabolism to synaptic deficits in autism spectrum disorders. Front Behav Neurosci. 2014;8(104):1–6.
79. Tint GS, Irons M, Elias ER, Batta AK, Frieden R, Chen TS, et al. Defective Cholesterol

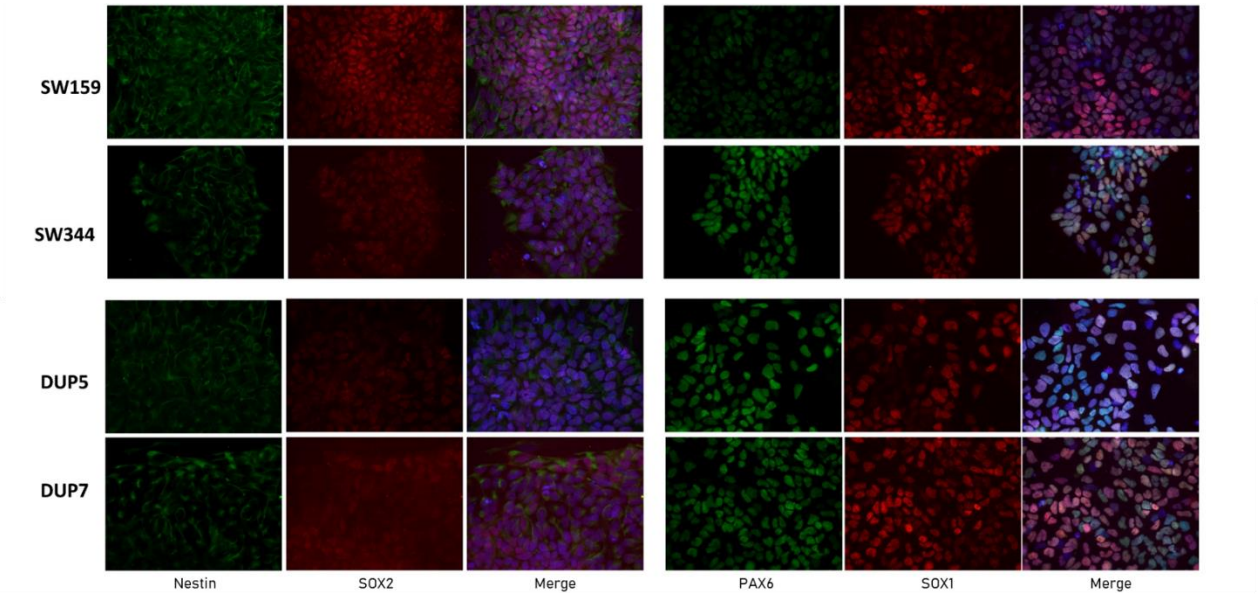
Biosynthesis Associated with the Smith-Lemli-Opitz Syndrome. N Engl J Med. 1994;330(2):107–13.
80. Tierney E, Nwokoro NA, Porter FD, Freund LS, Ghuman JK, Kelley RI. Behavior phenotype in the RSH/Smith-Lemli-Opitz syndrome. Am J Med Genet. 2001;98(2):191–200.
81. Sikora DM, Pettit-Kekel K, Penfield J, Merkens LS, Steiner RD. The near universal presence of autism spectrum disorders in children with Smith-Lemli-Opitz syndrome. Am J Med Genet Part A. 2006;140(14):1511–8.
82. Betancur C. Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Res. 2011;1380:42–77.
83. 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;52(4):248–55.
84. Ellaway CJ, Ho G, Bettella E, Knapman A, Collins F, Hackett A, et al. 14q12 microdeletions excluding FOXG1 give rise to a congenital variant Rett syndrome-like phenotype. Eur J Hum Genet. 2013;21:522–7.
85. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180(3):1–17.
86. Choudhury PR, Lahiri S, Rajamma U. Glutamate mediated signaling in the pathophysiology of autism spectrum disorders. Pharmacol Biochem Behav. 2012;100(4):841–9.
87. Purcell AE, Jeon OH, Zimmerman AW, Blue ME, Pevsner J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. Neurology. 2001;57(9):1618–28.
88. Tarabeux J, Kebir O, Gauthier J, Hamdan FF, Xiong L, Piton A, et al. Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. Transl Psychiatry. 2011;1:1–7.
89. 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;18(10):1795–804.
90. Gai X, Xie H, Perin J, Takahashi N, Murphy K, Wenocur A, et al. Rare structural variation of synapse and neurotransmission genes in autism. Mol Psychiatry. 2012;17:402–11.
91. Karayannis T, Au E, Patel JC, Kruglikov I, Markx S, Delorme R, et al. Cntnap4/Caspr4 Differentially Contributes to GABAergic and Dopaminergic Synaptic Transmission Nature. 2014;511(7508):236–40.
92. 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. 2012;485(7397):246–50.
93. Missler M, Südhof TC. Neurexophilins form a conserved family of neuropeptide-like glycoproteins. J Neurosci. 1998;18(10):3630–8.
94. Salyakina D, Cukier HN, Lee JM, Sacharow S, Nations LD. Copy Number Variants in Extended Autism Spectrum Disorder Families Reveal Candidates Potentially Involved in Autism Risk. PLoS One. 2011;6(10):e26049.
95. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, et al. Refinement and Discovery of New Hotspots of Copy-Number Variation Associated with Autism Spectrum Disorder. Am J Hum Genet. 2013;92:221–37.
96. Lu P, Chen X, Feng Y, Zeng Q, Jiang C, Zhu X, et al. Integrated transcriptome analysis of human iPS cells derived from a fragile X syndrome patient during neuronal differentiation. Sci China Life Sci. 2016;59(11):1093–105.
97. Chapman NH, Nato AQ, Bernier R, Ankeman K, Sohi H, Munson J, et al. Whole exome sequencing in extended families with autism spectrum disorder implicates four candidate genes HHS Public Access. Hum Genet. 2015;134(10):1055–68.

SUPPLEMENTARY INFORMATION

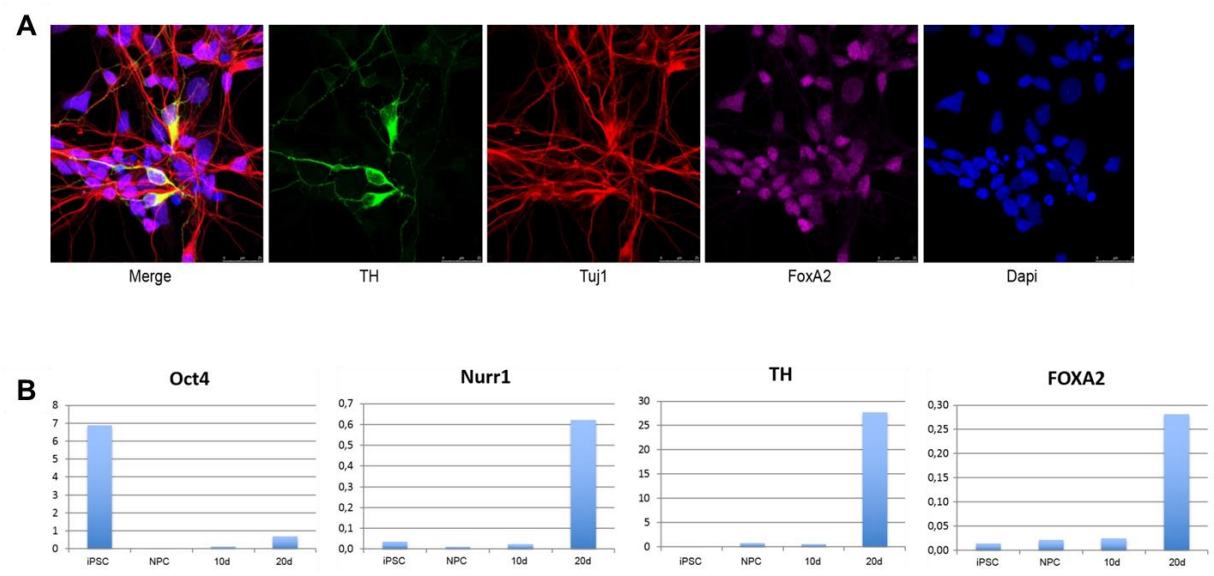
Supplementary Figure 1. (A) NPC induction and (B) dopaminergic neuron differentiation process.



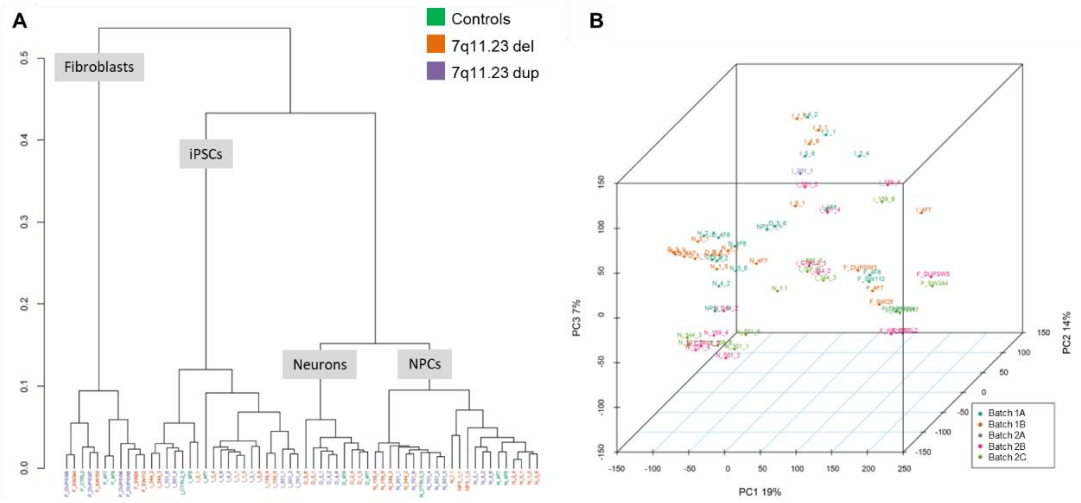
Supplementary Figure 2. Immunocytochemistry of 2 WBS NPC lines (patients SW159 and SW344) and 2 Dup7 NPC lines (patients DUPSW5 and DUPSW7).



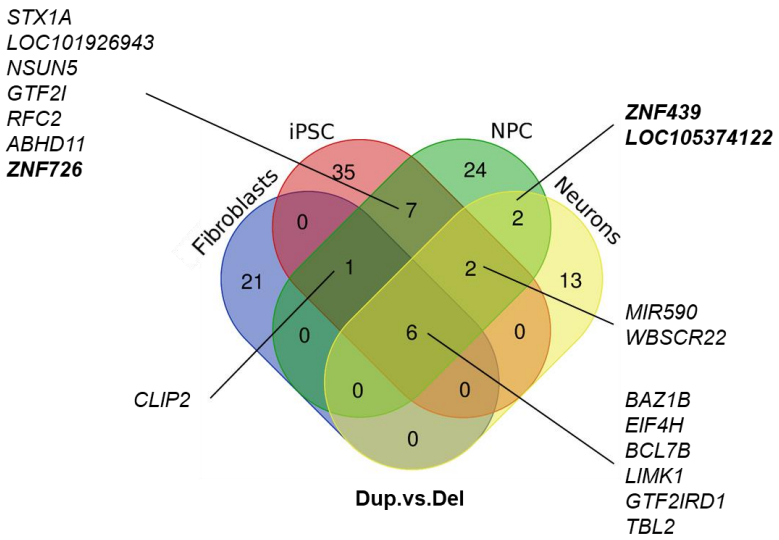
Supplementary Figure 3. (A) Immunocytochemistry and (B) qRT-PCR experiments using *GAPDH* as an internal control in Neurons (iNeus).



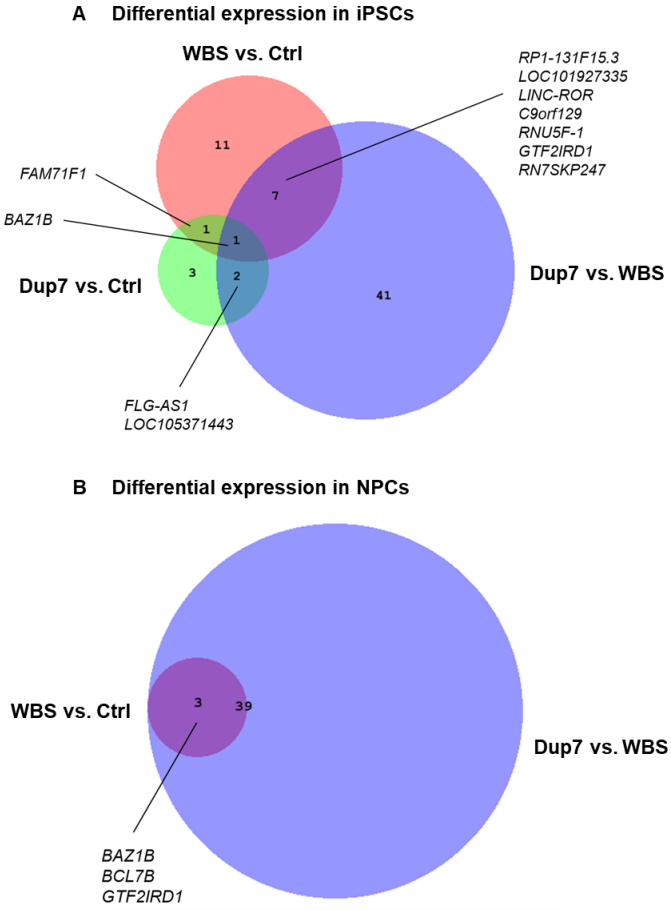
Supplementary Figure 4. (A) Dendogram and (B) PCA of samples included in the study.



Supplementary Figure 5. Venn diagram of the intersection between DEGs in the Dup.vs.Del comparison in all cell types. Bold indicates genes located outside the 7q11.23 interval.



Supplementary Figure 6. Overlap between DEGs in the three comparisons in (A) iPSCs and (B) NPCS.



Supplementary Table 1. Characteristics of subjects included in the study.

Samples	Skin biopsy	Sex	Age	7q11.23 CNV	Size (Mb)	Inversion	Origin	Associated neurodevelopmental features
SW26.01	2012	F	16	Del	1.55	No	Paternal	ID, hyperacusia
SW112.01	2012	F	16	Del	1.55	No	Maternal	ID, language delay, irritability
SW159.01 + Dup 16p13.11	2016	M	14	Del	1.55	Yes	Maternal	ID, ADHD
SW344.01	2015	M	7	Del	1.83	NA	Maternal	ID, ADHD symptomatology
DUPSW-3.01	2012	F	7	Dup	1.55	Yes	NA	ID
DUPSW-4.01	2012	F	14	Dup	1.55	Yes	NA	ID, language delay, anxiety, ADHD, OCD symptomatology
DUPSW-5.01	2012	M	17	Dup	2	No	Paternal	ASD, language delay, developmental delay
DUPSW-7.01	2015	M	7	Dup	1.55	Yes	Maternal	ASD, language delay, mild developmental delay, ADHD, anxiety
Control 1 4F7-4F8	-	M	2.5	No CNV	-	-	-	-
Control FiPS Ctrl2-R4F-5	-	M	3	No CNV	-	-	-	-

F: female; M: male; Del: deletion; Dup: duplication; Inter: interchromosomal; Intra: intrachromosomal; ID: intellectual disability; ADHD: attention deficit - hyperactivity disorder; OCD: obsessive-compulsive disorder; ASD: autism spectrum disorder

Supplementary Table 2. Primer sequences.

Gene	Primer F (5′-3′)	Primer R (5′-3′)
hGAPDH	GCACCGTCAAGGCTGAGAAC	AGGGATCTCGCTCCTGGAA
NURR1	CGAAGCCAGGGATCTTCTCT	GTGAGTCTGATCAGTGCCCT
TH	TCATCACCTGGTCACCAAGTT	GGTCGCCGTGCCTGTACT
FOXA2	TTCAGGCCCGGCTAACTCT	AGTCTCGACCCCCACTTGCT
D2S160	TGTACCTAAGCCCACCCTTTAGAGC	TGGCCTCCAGAAACCTCCAA
D5S406	CCTGCCAATACTTCAAGAAA	GGGATGCTAACTGCTGACTA
D8S260	AGGCTTGCCAGATAAGGTTG	GCTGAAGGCTGTTCTATGGA
D11S905	CAGGCATCTGAAACTTCTTG	ATACAGGGGCCAAATAGGTT
D13S158	GTACCCACGGAGTGAAAGAA	GCTTTGACAATTTAGCAGCA
D16S3091	GGGAGATAGCCTTAAACTTTCTTAC	TGTTGCTAATAACACTAGGCCA
D17S928	TAAAACGGCTACAACACATACA	ATTTCCCCACTGGCTG
D20S171	TATAGGTGAGGACCCTGAGG	ACACCAAGCCATGTAACCTG
D22S280	GCTCCAGCCTATCAGGATG	GATTCCAGATCACAAAACTGGT
CR16t	CTCTGGGAGTTCCCAAATGC	GAGGTTGCACTGAGCCAGA
WS5-ELN	GCCCACATGGGCAGATTGCT	CCCTCATCCACAGACAGGTC
D16S3060	TGAGCCGAGATTGCACCAA	TGCCACAGGGGATATAAGCC

Supplementary Table 3. Comparison between previous transcriptional studies of iPSC-based cells and derivatives and our data in 7q11.23 genes.

Gene_Symbol	Our_data Fibro-Del vs Ctrl	Henrichsen2011 Fibros-DelvsCtrl	Our_data iPS-Del vs Ctrl	Adamo2014 iPS-Del vs Ctrl	Lalli2016 iPS-Del vs Ctrl	Our_data iPS-Dup vs Ctrl	Adamo2014 iPS-Dup vs Ctrl	Our_data iNeu-Del vs Ctrl	Khattak2015 Neurons-Del vs Ctrl	Lalli2016 Neurons-Del vs Ctrl
NSUN5	Down*	Down**	Down	Down	Down	Up*	Up	Down	NA	Down*
TRIM50	Down	NA	Down	NA	NA	Down	NA	Up	NA	NA
FKBP6	Down	NA	Up	NA	NA	Down	NA	Up	NA	NA
FZD9	Up	NA	Down*	Down	Down*	Up	Up	Up	NA	Down*
BAZ1B	Down*	Down**	Down**	Down	Down	Up**	Up	Down**	Down	Down*
BCL7B	Down*	Down*	Down*	Down	Down	Up*	Up	Down*	Down	Down*
TBL2	Down*	Down**	Down*	Down	Down*	Up*	Up	Down*	Down	Down*
MLXIPL	Up	NA	Down*	Down	Down	Down	Down	Down	NA	NA
VPS37D	Down	NA	Down*	Down	Down*	Down	Up	Down	Down	Down*
DNAJC30	Down*	Down**	Down*	Down	NA	Up	Up	Down	Down	Down*
WBSCR22	Down*	Down**	Down*	Down	Down*	Up*	Up	Down*	Down	Down*
STX1A	Down	NA	Down	Down	Down	Up*	Up	Down*	Down	Down*
ABHD11-AS1	Down	NA	Down	NA	NA	Down	NA	Up	NA	NA
ABHD11	Down	NA	Down*	Down	Down	Up	Up	Down	NA	NA
CLDN3	Up	NA	Down*	Up	Down	Up	Up	Up	NA	NA
CLDN4	Down	NA	Down	Down	NA	Up*	Up	Up	NA	NA
WBSCR27	Down	NA	Down	Down	Down*	Up*	Up	Down	NA	NA
WBSCR28	Down	NA	Down	NA	NA	Down	NA	Down	NA	NA
ELN	Down	Down**	Down	Same	NA	Up	Up	Up	NA	NA
LIMK1	Down*	NA	Down*	Down	Down	Up*	Up	Down*	Down	Down*
EIF4H	Down*	Down**	Down*	Down	Down*	Up*	Up	Down*	Down	Down*
MIR590	Down*	NA	Down*	NA	NA	Up	NA	Down*	NA	NA
LAT2	Down	Up	Down	NA	NA	Down	NA	Down	NA	NA
RFC2	Down	Down	Down	Down	Down	Up*	Up	Down	Down	Down*
CLIP2	Down	Down	Down*	Down	Down	Up	Up	Down*	Down	Down*
GTF2IRD1	Down**	Down	Down**	Down	Down	Up*	Up	Down**	Down	Down*
GTF2I	Down*	Down**	Down*	Down	Down	Up	Up	Down*	Down	NA
LOC10192694	Down*	NA	Down*	NA	NA	Up*	NA	Down	NA	NA
NCF1	Down	NA	Down*	NA	NA	Down*	NA	Up	NA	NA
GTF2IRD2	Up	NA	Down	NA	NA	Down	NA	Down	Up	NA

First row colors: orange indicates columns related to fibroblasts, blue for iPSCs and yellow for neurons. Green text indicates a change in expression in the same direction, whereas red means it was found in the opposite direction. *: p-value<0.05, **: q-value<0.05.

Supplementary Table 4. Genome-wide differentially expressed genes among all cell types and comparisons.

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
17094893	9	72900662	73080442	*ALDH1A1*	-6.33	4.26E-28	1.57E-23	Fibro_Del vs Fibro_Ctrl
17015324	6	5997999	6007400	*NRN1*	4.79	5.15E-18	9.51E-14	Fibro_Del vs Fibro_Ctrl
16735764	11	10573091	10693988	*MRVI1*	-2.23	1.14E-16	1.40E-12	Fibro_Del vs Fibro_Ctrl
16851486	18	23689598	23955066	*LAMA3*	-1.87	4.91E-13	3.62E-09	Fibro_Del vs Fibro_Ctrl
16761212	12	9852369	9870136	*CLEC2B*	2.92	4.39E-13	3.62E-09	Fibro_Del vs Fibro_Ctrl
17078134	8	71197433	71362232	*EYA1*	-2.67	7.69E-13	4.73E-09	Fibro_Del vs Fibro_Ctrl
16978354	4	100395341	100880126	*EMCN*	1.72	1.19E-12	6.28E-09	Fibro_Del vs Fibro_Ctrl
17101214	X	2752050	2816500	*XG*	2.31	4.91E-12	2.27E-08	Fibro_Del vs Fibro_Ctrl
16945682	3	132317367	132368302	*ACPP*	-2.03	4.04E-11	1.36E-07	Fibro_Del vs Fibro_Ctrl
16736883	11	27696312	27877648	*RP11-587D21.4*	-2.01	4.03E-11	1.36E-07	Fibro_Del vs Fibro_Ctrl
16765450	12	53956375	53974956	*HOTAIR*	-1.69	3.35E-11	1.36E-07	Fibro_Del vs Fibro_Ctrl
16751917	12	53985065	53990279	*HOXC10*	-1.63	9.03E-11	2.60E-07	Fibro_Del vs Fibro_Ctrl
16671049	1	152698364	152699442	*LCE2A*	-1.55	9.17E-11	2.60E-07	Fibro_Del vs Fibro_Ctrl
17067748	8	32996178	33044855	*LOC105379362*	2.38	1.41E-10	3.72E-07	Fibro_Del vs Fibro_Ctrl
16754269	12	72087266	72665642	*TRHDE*	-3.59	3.07E-10	7.55E-07	Fibro_Del vs Fibro_Ctrl
16894824	2	19351485	19358653	*OSR1*	2.41	3.75E-10	8.67E-07	Fibro_Del vs Fibro_Ctrl
16975971	4	53459301	53701405	*LNX1*	1.50	4.31E-10	9.37E-07	Fibro_Del vs Fibro_Ctrl
16815791	16	10386055	10483638	*ATF7IP2*	-2.45	5.22E-10	1.02E-06	Fibro_Del vs Fibro_Ctrl
16668505	1	110082230	110110239	*LINC01397*	-2.15	5.52E-10	1.02E-06	Fibro_Del vs Fibro_Ctrl
16762724	12	29412474	29497686	*OVCH1*	-1.54	5.43E-10	1.02E-06	Fibro_Del vs Fibro_Ctrl
16979339	4	119494395	119628991	*PDE5A*	-2.71	7.74E-10	1.30E-06	Fibro_Del vs Fibro_Ctrl
16748529	12	12890782	12917937	*GPRC5A*	-2.29	7.47E-10	1.30E-06	Fibro_Del vs Fibro_Ctrl
17010273	6	72630495	72678558	*KCNQ5-IT1*	-1.77	8.80E-10	1.41E-06	Fibro_Del vs Fibro_Ctrl
17115271	X	153861514	153886173	*L1CAM*	-2.44	1.37E-09	2.10E-06	Fibro_Del vs Fibro_Ctrl
16821467	16	84294646	84329851	*WFDC1*	-2.54	3.00E-09	4.44E-06	Fibro_Del vs Fibro_Ctrl
16743721	11	102789910	102798235	*MMP1*	3.07	3.75E-09	5.33E-06	Fibro_Del vs Fibro_Ctrl
17104313	X	67544032	67730876	*AR*	-1.76	4.39E-09	6.01E-06	Fibro_Del vs Fibro_Ctrl
16697971	1	201896452	201946588	*LMOD1*	-1.82	6.67E-09	8.79E-06	Fibro_Del vs Fibro_Ctrl
16958251	3	123610049	123884331	*MYLK*	-1.67	8.90E-09	1.13E-05	Fibro_Del vs Fibro_Ctrl
16829728	17	3474110	3503419	*ASPA*	2.13	9.16E-09	1.13E-05	Fibro_Del vs Fibro_Ctrl
17094498	9	40992261	41074625	*PGM5P2*	-1.77	1.21E-08	1.44E-05	Fibro_Del vs Fibro_Ctrl
16983652	5	31193655	31329146	*CDH6*	-3.76	1.59E-08	1.84E-05	Fibro_Del vs Fibro_Ctrl
16987531	5	96875939	96919716	*ERAP2*	1.79	1.68E-08	1.88E-05	Fibro_Del vs Fibro_Ctrl
16852858	18	63752935	63805370	*SERPINB7*	-2.31	1.84E-08	1.89E-05	Fibro_Del vs Fibro_Ctrl
16711299	10	4200637	4243789	*LINC00702*	1.27	1.75E-08	1.89E-05	Fibro_Del vs Fibro_Ctrl
17012859	6	135851696	136195574	*PDE7B*	1.48	1.82E-08	1.89E-05	Fibro_Del vs Fibro_Ctrl
16836753	17	61452418	61485108	*TBX4*	-1.19	1.95E-08	1.95E-05	Fibro_Del vs Fibro_Ctrl
16914084	20	44714844	44728509	*WISP2*	2.04	2.57E-08	2.50E-05	Fibro_Del vs Fibro_Ctrl
16667516	1	99263953	99309590	*PLPPR4*	-2.67	3.01E-08	2.78E-05	Fibro_Del vs Fibro_Ctrl
16736861	11	27654893	27722058	*BDNF*	-1.44	2.94E-08	2.78E-05	Fibro_Del vs Fibro_Ctrl
16761221	12	9898673	9932381	*CLEC2A*	3.33	3.47E-08	3.13E-05	Fibro_Del vs Fibro_Ctrl
16721884	11	9758292	9811319	*SBF2-AS1*	-1.60	3.80E-08	3.27E-05	Fibro_Del vs Fibro_Ctrl
16998682	5	107376889	107670895	*EFNA5*	1.85	3.80E-08	3.27E-05	Fibro_Del vs Fibro_Ctrl
17077388	8	56535401	56540497	*RP11-17A4.3*	-2.09	4.03E-08	3.33E-05	Fibro_Del vs Fibro_Ctrl
16751910	12	53973126	53977643	*HOXC11*	-1.31	4.06E-08	3.33E-05	Fibro_Del vs Fibro_Ctrl
17022980	6	116752197	116765723	*FAM162B*	-1.38	4.17E-08	3.35E-05	Fibro_Del vs Fibro_Ctrl
17012761	6	133889121	133895553	*TCF21*	-1.85	6.18E-08	4.86E-05	Fibro_Del vs Fibro_Ctrl
16972446	4	173527270	173591324	*HAND2-AS1*	1.83	6.61E-08	5.09E-05	Fibro_Del vs Fibro_Ctrl
16919242	20	40685848	40689240	*MAFB*	1.62	7.22E-08	5.44E-05	Fibro_Del vs Fibro_Ctrl
16761293	12	10158300	10172191	*OLR1*	-1,55	1.01E-07	7.43E-05	Fibro_Del vs Fibro_Ctrl
16970118	4	118888841	119061247	*SYNPO2*	-2.12	1.05E-07	7.58E-05	Fibro_Del vs Fibro_Ctrl
16904488	2	164896186	164955525	*SLC38A11*	-0.92	1.07E-07	7.60E-05	Fibro_Del vs Fibro_Ctrl
16965142	4	14111968	14140052	*LINC01085*	-1.45	1.12E-07	7.79E-05	Fibro_Del vs Fibro_Ctrl
16767606	12	72253507	72274907	*TRHDE-AS1*	-2.18	1.16E-07	7.92E-05	Fibro_Del vs Fibro_Ctrl
16773681	13	30906175	30925572	*MEDAG*	2.48	1.23E-07	8.13E-05	Fibro_Del vs Fibro_Ctrl
16924645	21	28444549	28674848	*LOC102724420*	-0.97	1.23E-07	8.13E-05	Fibro_Del vs Fibro_Ctrl
16972249	4	168497066	168928457	*PALLD*	-1.37	1.41E-07	8.99E-05	Fibro_Del vs Fibro_Ctrl
16964888	4	8269762	8307111	*HTRA3*	1.36	1.41E-07	8.99E-05	Fibro_Del vs Fibro_Ctrl
16984294	5	40841184	40860173	*CARD6*	1.26	1.50E-07	9.38E-05	Fibro_Del vs Fibro_Ctrl
17052060	7	136868652	137020255	*M2*	-2.25	1.69E-07	0.00010	Fibro_Del vs Fibro_Ctrl
16872871	19	42866461	42879719	*PSG1*	2.38	1.74E-07	0.00011	Fibro_Del vs Fibro_Ctrl
17045897	7	47795291	47819847	*C7orf69*	-1.84	1.80E-07	0.00011	Fibro_Del vs Fibro_Ctrl
17094889	9	72871728	72874109	*LINC01474*	-1.62	2.38E-07	0.00014	Fibro_Del vs Fibro_Ctrl
16814738	16	1534230	1555242	*TMEM204*	1.33	2.42E-07	0.00014	Fibro_Del vs Fibro_Ctrl
16677451	1	215005542	215237093	*KCNK2*	1.74	2.46E-07	0.00014	Fibro_Del vs Fibro_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16696120	1	168695457	168729264	*DPT*	3.20	2.51E-07	0.00014	Fibro_Del vs Fibro_Ctrl
16821367	16	82035253	82098534	*HSD17B2*	-1.43	2.89E-07	0.00016	Fibro_Del vs Fibro_Ctrl
16679596	1	245154985	245703126	*KIF26B*	-1.70	3.40E-07	0.00018	Fibro_Del vs Fibro_Ctrl
16791371	14	24336021	24340045	*RIPK3*	1.20	4.39E-07	0.00023	Fibro_Del vs Fibro_Ctrl
17002667	5	169863715	169980740	*FAM196B*	1.35	4.38E-07	0.00023	Fibro_Del vs Fibro_Ctrl
16848219	17	69078706	69141874	*ABCA6*	2.07	4.69E-07	0.00024	Fibro_Del vs Fibro_Ctrl
16870453	19	18782773	18791314	*COMP*	1.36	5.33E-07	0.00027	Fibro_Del vs Fibro_Ctrl
17023782	6	133435077	133511728	*TARID*	-0.87	5.69E-07	0.00029	Fibro_Del vs Fibro_Ctrl
16945768	3	132597237	132618967	*ACKR4*	2.31	5.78E-07	0.00029	Fibro_Del vs Fibro_Ctrl
17078254	8	73064540	73124088	*SBSPON*	-1.11	5.97E-07	0.00029	Fibro_Del vs Fibro_Ctrl
16922114	21	31873315	32044633	*HUNK*	1.81	6.99E-07	0.00034	Fibro_Del vs Fibro_Ctrl
16687737	1	56494747	56645301	*PLPP3*	1.48	7.47E-07	0.00035	Fibro_Del vs Fibro_Ctrl
16977286	4	81088683	81215064	*PRKG2*	2.27	7.43E-07	0.00035	Fibro_Del vs Fibro_Ctrl
16995177	5	35048743	35230721	*PRLR*	2.33	7.34E-07	0.00035	Fibro_Del vs Fibro_Ctrl
17059388	7	86876906	87059699	*KIAA1324L*	2.27	8.52E-07	0.00039	Fibro_Del vs Fibro_Ctrl
16768738	12	95657807	95791152	*NTN4*	-1.93	1.03E-06	0.00047	Fibro_Del vs Fibro_Ctrl
16795819	14	91232532	91253925	*GPR68*	1.08	1.17E-06	0.00053	Fibro_Del vs Fibro_Ctrl
16755712	12	100717526	101128641	*ANO4*	-2.10	1.22E-06	0.00053	Fibro_Del vs Fibro_Ctrl
16994597	5	16473038	16617058	*FAM134B*	1.80	1.20E-06	0.00053	Fibro_Del vs Fibro_Ctrl
16907912	2	212999684	213152427	*IKZF2*	-0.93	1.22E-06	0.00053	Fibro_Del vs Fibro_Ctrl
17051553	7	130293134	130324180	*CPA4*	-1.70	1.57E-06	0.00067	Fibro_Del vs Fibro_Ctrl
17095887	9	92456205	92482562	*ASPN*	-1.63	1.58E-06	0.00067	Fibro_Del vs Fibro_Ctrl
17055828	7	23310209	23470467	*IGF2BP3*	-1.68	1.69E-06	0.00071	Fibro_Del vs Fibro_Ctrl
17014091	6	157823249	157945077	*SNX9*	1.19	1.80E-06	0.00075	Fibro_Del vs Fibro_Ctrl
16960911	3	158645822	158672693	*LXN*	-1.80	1.84E-06	0.00075	Fibro_Del vs Fibro_Ctrl
17019698	6	46220730	46492067	*RCAN2*	2.37	1.86E-06	0.00076	Fibro_Del vs Fibro_Ctrl
16707503	10	92831178	93059494	*EXOC6*	-1.43	1.93E-06	0.00077	Fibro_Del vs Fibro_Ctrl
16713052	10	31693084	31707388	*LOC101929951*	1.44	1.98E-06	0.00079	Fibro_Del vs Fibro_Ctrl
16948168	3	177683627	177691234	*RP11-114M1.1*	-1.19	2.01E-06	0.00079	Fibro_Del vs Fibro_Ctrl
16914925	20	52972407	53495330	*TSHZ2*	1.50	2.07E-06	0.00080	Fibro_Del vs Fibro_Ctrl
16780929	13	110148963	110307149	*COL4A1*	-1.86	2.48E-06	0.00095	Fibro_Del vs Fibro_Ctrl
16732394	11	120213736	120217513	*OAF*	-1.03	2.59E-06	0.00098	Fibro_Del vs Fibro_Ctrl
16908138	2	215942591	216034096	*MREG*	1.26	2.59E-06	0.00098	Fibro_Del vs Fibro_Ctrl
16712168	10	16823964	17129831	*CUBN*	1.43	2.90E-06	0.00108	Fibro_Del vs Fibro_Ctrl
16961331	3	169083499	169663775	*MECOM*	-1.58	3.02E-06	0.00112	Fibro_Del vs Fibro_Ctrl
16835517	17	48997412	49056145	*IGF2BP1*	-1.04	3.12E-06	0.00114	Fibro_Del vs Fibro_Ctrl
17077376	8	56494282	56559823	*LINC00968*	-1.52	3.24E-06	0.00116	Fibro_Del vs Fibro_Ctrl
16981706	4	176213673	176277571	*ASB5*	-1.03	3.25E-06	0.00116	Fibro_Del vs Fibro_Ctrl
16998293	5	95751319	95823005	*GLRX*	-1.01	3.46E-06	0.00123	Fibro_Del vs Fibro_Ctrl
16690704	1	110362848	110391082	*SLC16A4*	-1.36	3.70E-06	0.00130	Fibro_Del vs Fibro_Ctrl
17063221	7	135728348	135748846	*FAM180A*	1.12	4.03E-06	0.00139	Fibro_Del vs Fibro_Ctrl
17021596	6	89364616	89412276	*RRAGD*	1.86	4.04E-06	0.00139	Fibro_Del vs Fibro_Ctrl
16999475	5	128257909	128659185	*FBN2*	-3.49	4.09E-06	0.00140	Fibro_Del vs Fibro_Ctrl
16952912	3	46515386	46580099	*LRRC2*	1.61	4.33E-06	0.00147	Fibro_Del vs Fibro_Ctrl
16718622	10	114431110	114684655	*ABLIM1*	-1.04	4.41E-06	0.00148	Fibro_Del vs Fibro_Ctrl
17074547	8	10482878	10547585	*PRSS51*	1.33	4.80E-06	0.00160	Fibro_Del vs Fibro_Ctrl
16980762	4	153780590	153789120	*SFRP2*	2.71	5.14E-06	0.00168	Fibro_Del vs Fibro_Ctrl
17115375	X	153935263	153944780	*RENBP*	0.97	5.14E-06	0.00168	Fibro_Del vs Fibro_Ctrl
16691576	1	118883043	118989556	*TBX15*	1.52	5.48E-06	0.00177	Fibro_Del vs Fibro_Ctrl
17023799	6	133987581	134052651	*SLC2A12*	-1.95	5.87E-06	0.00188	Fibro_Del vs Fibro_Ctrl
16999557	5	128371099	128371212	*Y_RNA/FBN2*	-2.34	5.98E-06	0.00190	Fibro_Del vs Fibro_Ctrl
16792954	14	52004802	52069827	*NID2*	-1.98	6.06E-06	0.00191	Fibro_Del vs Fibro_Ctrl
16962536	3	187216085	187292022	*MASP1*	-1.30	6.10E-06	0.00191	Fibro_Del vs Fibro_Ctrl
16706456	10	78352597	78356100	*LOC105378377*	1.18	6.15E-06	0.00191	Fibro_Del vs Fibro_Ctrl
16946707	3	148697784	148743164	*AGTR1*	0.88	6.73E-06	0.00207	Fibro_Del vs Fibro_Ctrl
16922759	21	38157034	38301826	*KCNJ15*	1.01	9.38E-06	0.00284	Fibro_Del vs Fibro_Ctrl
17012047	6	117453817	117569858	*DCBLD1*	-1.19	9.96E-06	0.00297	Fibro_Del vs Fibro_Ctrl
16668220	1	109466478	109482142	*SYPL2*	-1.18	9.96E-06	0.00297	Fibro_Del vs Fibro_Ctrl
17044177	7	22725884	22732002	*IL6*	-1.13	1.02E-05	0.00303	Fibro_Del vs Fibro_Ctrl
16811512	15	73739800	73752378	*C15orf59*	1.16	1.07E-05	0.00310	Fibro_Del vs Fibro_Ctrl
16701037	1	240489573	240612162	*GREM2*	1.75	1.07E-05	0.00310	Fibro_Del vs Fibro_Ctrl
16769552	12	105304867	105327017	*KCCAT198*	1.64	1.08E-05	0.00312	Fibro_Del vs Fibro_Ctrl
16990897	5	149271838	149341805	*AFAP1L1*	0.78	1.11E-05	0.00317	Fibro_Del vs Fibro_Ctrl
17068349	8	40153468	40155308	*C8orf4*	-2.08	1.14E-05	0.00324	Fibro_Del vs Fibro_Ctrl
16997383	5	76615482	76623434	*F2RL2*	-2.17	1.20E-05	0.00335	Fibro_Del vs Fibro_Ctrl
16821654	16	86510527	86514470	*FOXF1*	-1.47	1.21E-05	0.00335	Fibro_Del vs Fibro_Ctrl
17078539	8	80628451	80874781	*ZNF704*	1.65	1.20E-05	0.00335	Fibro_Del vs Fibro_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16948161	3	177441921	177752704	*LINC00578*	-0.99	1.23E-05	0.00339	Fibro_Del vs Fibro_Ctrl
16775491	13	75871038	75883812	*LMO7DN*	-1.19	1.37E-05	0.00375	Fibro_Del vs Fibro_Ctrl
16804506	15	88803443	88875354	*ACAN*	-1.66	1.38E-05	0.00376	Fibro_Del vs Fibro_Ctrl
17017641	6	32041154	32115334	*TNXB*	2.01	1.40E-05	0.00377	Fibro_Del vs Fibro_Ctrl
16675197	1	186828900	186988981	*PLA2G4A*	2.42	1.48E-05	0.00397	Fibro_Del vs Fibro_Ctrl
16832227	17	20579724	20580911	*CDRT15L2*	-1.18	1.52E-05	0.00404	Fibro_Del vs Fibro_Ctrl
16818114	16	30985198	30989152	*HSD3B7*	1.03	1.64E-05	0.00433	Fibro_Del vs Fibro_Ctrl
16703036	10	19816239	20280357	*PLXDC2*	2.45	1.82E-05	0.00473	Fibro_Del vs Fibro_Ctrl
16852789	18	62715439	62980443	*PHLPP1*	0.94	1.82E-05	0.00473	Fibro_Del vs Fibro_Ctrl
17061881	7	111726108	112206411	*DOCK4*	1.90	1.89E-05	0.00489	Fibro_Del vs Fibro_Ctrl
16839277	17	799313	979770	*NXN*	-1.05	2.03E-05	0.00522	Fibro_Del vs Fibro_Ctrl
16829085	16	87830016	87869519	*SLC7A5*	-1.87	2.14E-05	0.00545	Fibro_Del vs Fibro_Ctrl
16679333	1	240530452	240530862	*RP11-467I20.2*	1.70	2.17E-05	0.00549	Fibro_Del vs Fibro_Ctrl
16964597	4	5051442	5500998	*STK32B*	1.44	2.22E-05	0.00557	Fibro_Del vs Fibro_Ctrl
16718779	10	117473215	117545068	*EMX2OS*	1.36	2.23E-05	0.00557	Fibro_Del vs Fibro_Ctrl
16888437	2	184598366	184939492	*ZNF804A*	-1.43	2.63E-05	0.00652	Fibro_Del vs Fibro_Ctrl
16759262	12	130977562	130978768	*RP11-243M5.1*	-1.71	2.70E-05	0.00659	Fibro_Del vs Fibro_Ctrl
16935905	22	44243667	44312851	*KIAA1644*	1.63	2.68E-05	0.00659	Fibro_Del vs Fibro_Ctrl
16901986	2	112829751	112836903	*IL1B*	-0.87	2.78E-05	0.00674	Fibro_Del vs Fibro_Ctrl
16831525	17	16783623	16783733	*RP11-219A15.1*	-2.23	2.96E-05	0.00714	Fibro_Del vs Fibro_Ctrl
16904667	2	166195185	166375993	*SCN9A*	-1.78	3.01E-05	0.00715	Fibro_Del vs Fibro_Ctrl
16844775	17	41523617	41528389	*KRT19*	-1.44	3.02E-05	0.00715	Fibro_Del vs Fibro_Ctrl
17107804	X	151695124	151705924	*PRRG3*	0.99	3.01E-05	0.00715	Fibro_Del vs Fibro_Ctrl
17115782	X	155276211	155334677	*CLIC2*	1.68	3.08E-05	0.00724	Fibro_Del vs Fibro_Ctrl
17076244	8	37516450	37517021	*LINC01605*	1.01	3.13E-05	0.00732	Fibro_Del vs Fibro_Ctrl
16843241	17	32492610	32876884	*MYO1D*	1.59	3.24E-05	0.00743	Fibro_Del vs Fibro_Ctrl
17085516	9	41073710	41076392	*PGM5-AS1*	-0.81	3.24E-05	0.00743	Fibro_Del vs Fibro_Ctrl
16706117	10	73772291	73779092	*FUT11*	0.77	3.23E-05	0.00743	Fibro_Del vs Fibro_Ctrl
16919158	20	38128461	38166578	*TGM2*	-1.52	3.29E-05	0.00749	Fibro_Del vs Fibro_Ctrl
16873826	19	48047843	48110852	*PLA2G4C*	0.86	3.35E-05	0.00759	Fibro_Del vs Fibro_Ctrl
16855673	18	63123346	63320128	*BCL2*	1.29	3.47E-05	0.00782	Fibro_Del vs Fibro_Ctrl
16925398	21	35037925	35039426	*RUNX1-IT1*	-1.25	3.60E-05	0.00807	Fibro_Del vs Fibro_Ctrl
16997799	5	83541477	83581320	*VCAN-AS1*	2.28	3.68E-05	0.00819	Fibro_Del vs Fibro_Ctrl
16951567	3	24117153	24495282	*THRB*	1.43	3.88E-05	0.00857	Fibro_Del vs Fibro_Ctrl
16671065	1	152776372	152776969	*LCE1F*	-1.17	3.95E-05	0.00868	Fibro_Del vs Fibro_Ctrl
16762759	12	29500813	29784759	*TMTC1*	-1.87	4.24E-05	0.00922	Fibro_Del vs Fibro_Ctrl
17056190	7	27187247	27192686	*RP1-170O19.14*	-1.01	4.25E-05	0.00922	Fibro_Del vs Fibro_Ctrl
16972291	4	168885355	168885461	*RNU6-853P*	-3.06	4.64E-05	0.01002	Fibro_Del vs Fibro_Ctrl
16775498	13	75878713	75881127	*LMO7DN-IT1*	-0.96	4.71E-05	0.01011	Fibro_Del vs Fibro_Ctrl
16759218	12	130953907	131141469	*ADGRD1*	-1.52	5.13E-05	0.01085	Fibro_Del vs Fibro_Ctrl
16765464	12	53983450	53985519	*HOXC-AS3*	-1.00	5.14E-05	0.01085	Fibro_Del vs Fibro_Ctrl
16716371	10	89205937	89207314	*CH25H*	1.93	5.12E-05	0.01085	Fibro_Del vs Fibro_Ctrl
17070013	8	73294602	73325285	*RDH10*	-1.07	5.34E-05	0.01121	Fibro_Del vs Fibro_Ctrl
16689354	1	89107627	89126116	*GBP2*	1.03	5.44E-05	0.01134	Fibro_Del vs Fibro_Ctrl
16745987	11	126920188	126940659	*LOC105369559*	-2.02	5.48E-05	0.01137	Fibro_Del vs Fibro_Ctrl
16690211	1	102876467	103108496	*COL11A1*	-2.26	5.53E-05	0.01141	Fibro_Del vs Fibro_Ctrl
16971382	4	150078274	150257931	*DCLK2*	-1.05	5.64E-05	0.01157	Fibro_Del vs Fibro_Ctrl
16738477	11	57399496	57399566	*PRG2*	1.63	5.67E-05	0.01157	Fibro_Del vs Fibro_Ctrl
16924602	21	26836287	26845409	*ADAMTS1*	-1.71	5.78E-05	0.01174	Fibro_Del vs Fibro_Ctrl
16977986	4	89724099	89838315	*SNCA*	-1.02	5.92E-05	0.01195	Fibro_Del vs Fibro_Ctrl
17010246	6	72621792	73198851	*KCNQ5*	-1.28	6.00E-05	0.01199	Fibro_Del vs Fibro_Ctrl
16937741	3	12287368	12434356	*PPARG*	-1.10	6.02E-05	0.01199	Fibro_Del vs Fibro_Ctrl
16801487	15	58410569	58568952	*LIPC*	1.10	6.33E-05	0.01251	Fibro_Del vs Fibro_Ctrl
16796694	14	100333788	100376805	*WARS*	-0.96	6.48E-05	0.01272	Fibro_Del vs Fibro_Ctrl
16797196	14	104937244	104978357	*AHNAK2*	1.03	6.63E-05	0.01295	Fibro_Del vs Fibro_Ctrl
16694689	1	156668763	156677397	*NES*	-1.67	6.79E-05	0.01298	Fibro_Del vs Fibro_Ctrl
16690682	1	110059994	110070700	*ALX3*	-1.06	6.78E-05	0.01298	Fibro_Del vs Fibro_Ctrl
16775643	13	79481124	79556077	*NDFIP2*	-0.92	6.75E-05	0.01298	Fibro_Del vs Fibro_Ctrl
16865112	19	53864763	53866140	*MYADM*	0.95	6.68E-05	0.01298	Fibro_Del vs Fibro_Ctrl
16832147	17	19648146	19677596	*ALDH3A2*	1.08	7.13E-05	0.01357	Fibro_Del vs Fibro_Ctrl
16802653	15	71096952	71783383	*THSD4*	-1.70	7.37E-05	0.01396	Fibro_Del vs Fibro_Ctrl
16831529	17	16786489	16816540	*USP32P1*	-1.52	7.46E-05	0.01406	Fibro_Del vs Fibro_Ctrl
16911418	20	10218829	10307418	*SNAP25*	-1.26	7.87E-05	0.01474	Fibro_Del vs Fibro_Ctrl
16778067	13	35768652	36131377	*DCLK1*	1.47	8.02E-05	0.01492	Fibro_Del vs Fibro_Ctrl
16754880	12	85280107	85301784	*ALX1*	1.63	8.04E-05	0.01492	Fibro_Del vs Fibro_Ctrl
17084127	9	27102630	27104728	*RP11-56F10.3*	-1.21	8.09E-05	0.01494	Fibro_Del vs Fibro_Ctrl
16723353	11	33015864	33033582	*DEPDC7*	0.86	8.55E-05	0.01563	Fibro_Del vs Fibro_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16801473	15	58138169	58185911	*AQP9*	0.97	8.51E-05	0.01563	Fibro_Del vs Fibro_Ctrl
16899357	2	74532414	74555690	*LOXL3*	-1.29	9.25E-05	0.01676	Fibro_Del vs Fibro_Ctrl
16906749	2	196199253	196593692	*HECW2*	0.96	9.26E-05	0.01676	Fibro_Del vs Fibro_Ctrl
16748327	12	10226058	10317251	*KLRD1*	-0.77	9.34E-05	0.01682	Fibro_Del vs Fibro_Ctrl
16757373	12	112978395	113011723	*OAS2*	0.87	9.45E-05	0.01694	Fibro_Del vs Fibro_Ctrl
17059451	7	87401697	87480435	*ABCB4*	0.89	0.00010	0.01712	Fibro_Del vs Fibro_Ctrl
16955197	3	55465715	55489945	*WNT5A*	-2.40	0.00010	0.01724	Fibro_Del vs Fibro_Ctrl
16998287	5	95701249	95732295	*CTD-2154I11.2*	-0.75	0.00010	0.01724	Fibro_Del vs Fibro_Ctrl
16973498	4	1166932	1208962	*SPON2*	0.77	0.00010	0.01724	Fibro_Del vs Fibro_Ctrl
16961413	3	169939353	169966734	*LOC100128164*	-0.75	0.00010	0.01772	Fibro_Del vs Fibro_Ctrl
16904105	2	159768628	159904756	*LY75-CD302*	1.60	0.00010	0.01811	Fibro_Del vs Fibro_Ctrl
17093204	9	33247820	33264763	*BAG1*	0.85	0.00011	0.01861	Fibro_Del vs Fibro_Ctrl
16811085	15	68301704	68432163	*ITGA11*	1.96	0.00011	0.01872	Fibro_Del vs Fibro_Ctrl
16880814	2	66433452	66572759	*MEIS1*	-1.43	0.00011	0.01956	Fibro_Del vs Fibro_Ctrl
16876777	2	6917392	7068286	*RNF144A*	1.06	0.00011	0.01960	Fibro_Del vs Fibro_Ctrl
17095499	9	86944362	86947189	*GAS1*	1.35	0.00012	0.02006	Fibro_Del vs Fibro_Ctrl
16802251	15	67063763	67195195	*SMAD3*	-0.91	0.00012	0.02006	Fibro_Del vs Fibro_Ctrl
16969817	4	112145397	112189081	*C4orf32*	1.11	0.00012	0.02015	Fibro_Del vs Fibro_Ctrl
16949827	3	195147871	195152790	*XXYLT1-AS2*	1.10	0.00012	0.02035	Fibro_Del vs Fibro_Ctrl
16854243	18	23136435	23136512	*AC105247.1*	1.11	0.00012	0.02049	Fibro_Del vs Fibro_Ctrl
17002194	5	157395534	157575775	*ADAM19*	-1.20	0.00013	0.02138	Fibro_Del vs Fibro_Ctrl
16986065	5	73120292	73131817	*TMEM171*	0.84	0.00013	0.02138	Fibro_Del vs Fibro_Ctrl
16853716	18	6941744	7117814	*LAMA1*	-1.92	0.00013	0.02170	Fibro_Del vs Fibro_Ctrl
16698883	1	209614870	209652475	*LAMB3*	-0.97	0.00013	0.02176	Fibro_Del vs Fibro_Ctrl
16967843	4	74365143	74388760	*EREG*	-1.12	0.00014	0.02218	Fibro_Del vs Fibro_Ctrl
16956792	3	100749156	100993515	*ABI3BP*	1.19	0.00014	0.02236	Fibro_Del vs Fibro_Ctrl
16972167	4	165361194	165498330	*CPE*	1.47	0.00014	0.02236	Fibro_Del vs Fibro_Ctrl
16950440	3	8750408	8769628	*OXTR*	-1.86	0.00014	0.02305	Fibro_Del vs Fibro_Ctrl
16689960	1	96254069	96374125	*LOC101928241*	0.69	0.00015	0.02329	Fibro_Del vs Fibro_Ctrl
16846218	17	48540894	48546079	*HOXB2*	1.59	0.00015	0.02346	Fibro_Del vs Fibro_Ctrl
16952734	3	44617128	44624797	*ZNF197-AS1*	-0.67	0.00016	0.02581	Fibro_Del vs Fibro_Ctrl
16809926	15	58856831	58865063	*RP11-30K9.5*	0.84	0.00017	0.02695	Fibro_Del vs Fibro_Ctrl
16852871	18	63871692	63903890	*SERPINB2*	-1.84	0.00017	0.02697	Fibro_Del vs Fibro_Ctrl
16922469	21	36005338	36007601	*LINC01436*	1.42	0.00018	0.02785	Fibro_Del vs Fibro_Ctrl
16960084	3	143265222	143848531	*SLC9A9*	1.23	0.00018	0.02802	Fibro_Del vs Fibro_Ctrl
16998097	5	93409359	93585648	*NR2F1-AS1*	-1.00	0.00018	0.02817	Fibro_Del vs Fibro_Ctrl
16813565	15	96127369	96327361	*NR2F2-AS1*	-1.38	0.00019	0.02905	Fibro_Del vs Fibro_Ctrl
17095375	9	85546539	85742029	*AGTPBP1*	0.91	0.00019	0.02955	Fibro_Del vs Fibro_Ctrl
16980946	4	156760454	156971394	*PDGFC*	1.20	0.00019	0.02984	Fibro_Del vs Fibro_Ctrl
16755173	12	94148723	94307675	*PLXNC1*	1.26	0.00020	0.03092	Fibro_Del vs Fibro_Ctrl
16908685	2	219236606	219245478	*GLB1L*	1.02	0.00020	0.03115	Fibro_Del vs Fibro_Ctrl
16657506	1	817371	819834	*FAM87B*	0.81	0.00021	0.03148	Fibro_Del vs Fibro_Ctrl
16803304	15	76336724	76342476	*ISL2*	-0.98	0.00022	0.03268	Fibro_Del vs Fibro_Ctrl
17044862	7	30911800	30925516	*AQP1*	1.67	0.00022	0.03300	Fibro_Del vs Fibro_Ctrl
17051626	7	130486175	130506493	*MEST*	-1.65	0.00022	0.03308	Fibro_Del vs Fibro_Ctrl
16666357	1	77281977	77559969	*AK5*	1.10	0.00023	0.03437	Fibro_Del vs Fibro_Ctrl
16986655	5	79991311	80083287	*THBS4*	2.58	0.00023	0.03498	Fibro_Del vs Fibro_Ctrl
16731297	11	112961247	113278436	*NCAM1*	1.22	0.00024	0.03519	Fibro_Del vs Fibro_Ctrl
16738630	11	58526871	58578220	*LPXN*	1.30	0.00024	0.03519	Fibro_Del vs Fibro_Ctrl
17099403	9	133097720	133149334	*RALGDS*	0.72	0.00024	0.03534	Fibro_Del vs Fibro_Ctrl
16980470	4	148078762	148444698	*NR3C2*	0.85	0.00024	0.03557	Fibro_Del vs Fibro_Ctrl
16754646	12	80405994	80680234	*PTPRQ*	-1.20	0.00025	0.03559	Fibro_Del vs Fibro_Ctrl
16658536	1	7784320	7845177	*PER3*	1.11	0.00025	0.03559	Fibro_Del vs Fibro_Ctrl
16905436	2	176629581	176637931	*LINC01116*	1.11	0.00025	0.03559	Fibro_Del vs Fibro_Ctrl
16889455	2	201182881	201229406	*CASP10*	0.59	0.00025	0.03559	Fibro_Del vs Fibro_Ctrl
17088809	9	122614669	122615747	*OR1Q1*	0.84	0.00025	0.03559	Fibro_Del vs Fibro_Ctrl
16956316	3	74262568	74521192	*CNTN3*	1.36	0.00026	0.03729	Fibro_Del vs Fibro_Ctrl
17059567	7	88270892	88306913	*STEAP4*	1.55	0.00026	0.03729	Fibro_Del vs Fibro_Ctrl
16733325	11	126940746	126945088	*KIRREL3-AS2*	-0.82	0.00027	0.03823	Fibro_Del vs Fibro_Ctrl
16703756	10	32056746	32056839	*Y_RNA*	1.10	0.00027	0.03858	Fibro_Del vs Fibro_Ctrl
16776431	13	110305812	110513027	*COL4A2*	-1.21	0.00027	0.03872	Fibro_Del vs Fibro_Ctrl
17072271	8	121639293	121645325	*HAS2-AS1*	0.62	0.00028	0.03930	Fibro_Del vs Fibro_Ctrl
17062127	7	117276631	117323289	*WNT2*	-1.07	0.00029	0.04038	Fibro_Del vs Fibro_Ctrl
17044414	7	26152227	26187137	*NFE2L3*	-1.63	0.00029	0.04062	Fibro_Del vs Fibro_Ctrl
16914972	20	54475597	54651171	*DOK5*	1.47	0.00030	0.04143	Fibro_Del vs Fibro_Ctrl
16668513	1	110112440	110113947	*UBL4B*	-0.95	0.00030	0.04158	Fibro_Del vs Fibro_Ctrl
16705188	10	61662961	61766766	*C10orf107*	-1.16	0.00030	0.04170	Fibro_Del vs Fibro_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16910000	2	238788648	238789716	*AC144525.1*	0.68	0.00031	0.04202	Fibro_Del vs Fibro_Ctrl
16919915	20	47209214	47356823	*ZMYND8*	-0.58	0.00031	0.04256	Fibro_Del vs Fibro_Ctrl
17096471	9	98198998	98255721	*TBC1D2*	-1.21	0.00031	0.04269	Fibro_Del vs Fibro_Ctrl
16730967	11	109422120	109429167	*C11orf87*	0.99	0.00032	0.04375	Fibro_Del vs Fibro_Ctrl
16668079	1	108875350	108930896	*GPSM2*	1.36	0.00032	0.04384	Fibro_Del vs Fibro_Ctrl
16833164	17	32876759	32878637	*LOC102724715*	0.83	0.00033	0.04471	Fibro_Del vs Fibro_Ctrl
16779839	13	76880169	76886405	*KCTD12*	1.53	0.00033	0.04479	Fibro_Del vs Fibro_Ctrl
17047138	7	74453790	74602605	*GTF2IRD1*	-0.57	0.00034	0.04547	Fibro_Del vs Fibro_Ctrl
16920475	20	53936777	54070765	*BCAS1*	0.80	0.00035	0.04602	Fibro_Del vs Fibro_Ctrl
17071769	8	105669991	105679031	*RP11-642D21.1*	-1.39	0.00035	0.04602	Fibro_Del vs Fibro_Ctrl
16686060	1	42925375	42959176	*SLC2A1*	-1.44	0.00035	0.04603	Fibro_Del vs Fibro_Ctrl
16908197	2	216672105	216695549	*IGFBP5*	-2.50	0.00036	0.04770	Fibro_Del vs Fibro_Ctrl
16744049	11	107790991	107928293	*SLC35F2*	0.81	0.00036	0.04782	Fibro_Del vs Fibro_Ctrl
16920315	20	51386957	51562800	*NFATC2*	1.14	0.00037	0.04810	Fibro_Del vs Fibro_Ctrl
16827564	16	68017710	68017817	*RNU6-359P*	1.01	0.00037	0.04832	Fibro_Del vs Fibro_Ctrl
16665439	1	62194802	62212328	*L1TD1*	1.12	0.00037	0.04832	Fibro_Del vs Fibro_Ctrl
16853460	18	797989	798135	*RNU1-109P*	0.73	0.00037	0.04832	Fibro_Del vs Fibro_Ctrl
16777715	13	28300346	28495128	*FLT1*	-1.29	0.00038	0.04851	Fibro_Del vs Fibro_Ctrl
16674927	1	183636073	183928532	*RGL1*	0.87	0.00038	0.04851	Fibro_Del vs Fibro_Ctrl
16914839	20	50301095	50302042	*LINC01270*	0.55	0.00038	0.04872	Fibro_Del vs Fibro_Ctrl
16716420	10	89579497	89645572	*PANK1*	0.65	0.00039	0.04964	Fibro_Del vs Fibro_Ctrl
16706445	10	78267328	78330595	*LINC00595*	0.81	0.00039	0.04964	Fibro_Del vs Fibro_Ctrl
17094893	9	72900662	73080442	*ALDH1A1*	-6.36	3.22E-28	1.19E-23	Fibro_Dup vs Fibro_Ctrl
16735764	11	10573091	10693988	*MRVI1*	-2.35	1.65E-17	3.05E-13	Fibro_Dup vs Fibro_Ctrl
17067748	8	32996178	33044855	*LOC105379362*	3.70	3.89E-17	4.47E-13	Fibro_Dup vs Fibro_Ctrl
17015324	6	5997999	6007400	*NRN1*	4.53	4.84E-17	4.47E-13	Fibro_Dup vs Fibro_Ctrl
16851486	18	23689598	23955066	*LAMA3*	-2.38	7.04E-17	5.20E-13	Fibro_Dup vs Fibro_Ctrl
16765450	12	53956375	53974956	*HOTAIR*	-1.98	2.13E-13	1.31E-09	Fibro_Dup vs Fibro_Ctrl
16761212	12	9852369	9870136	*CLEC2B*	2.62	1.43E-11	7.52E-08	Fibro_Dup vs Fibro_Ctrl
16945682	3	132317367	132368302	*ACPP*	-2.07	1.96E-11	9.05E-08	Fibro_Dup vs Fibro_Ctrl
17010273	6	72630495	72678558	*KCNQ5-IT1*	-1.94	6.17E-11	2.53E-07	Fibro_Dup vs Fibro_Ctrl
17078134	8	71197433	71362232	*EYA1*	-2.29	9.46E-11	3.49E-07	Fibro_Dup vs Fibro_Ctrl
16671049	1	152698364	152699442	*LCE2A*	-1.48	3.00E-10	1.01E-06	Fibro_Dup vs Fibro_Ctrl
17101214	X	2752050	2816500	*XG*	1.99	4.47E-10	1.38E-06	Fibro_Dup vs Fibro_Ctrl
16919242	20	40685848	40689240	*MAFB*	1.96	6.53E-10	1.86E-06	Fibro_Dup vs Fibro_Ctrl
16814738	16	1534230	1555242	*TMEM204*	1.67	8.73E-10	2.30E-06	Fibro_Dup vs Fibro_Ctrl
16748529	12	12890782	12917937	*GPRC5A*	-2.25	1.19E-09	2.92E-06	Fibro_Dup vs Fibro_Ctrl
16904488	2	164896186	164955525	*SLC38A11*	-1.10	1.35E-09	3.11E-06	Fibro_Dup vs Fibro_Ctrl
16924645	21	28444549	28674848	*LOC102724420*	-1.15	1.83E-09	3.99E-06	Fibro_Dup vs Fibro_Ctrl
17077388	8	56535401	56540497	*RP11-17A4.3*	-2.34	2.30E-09	4.71E-06	Fibro_Dup vs Fibro_Ctrl
16762724	12	29412474	29497686	*OVCH1*	-1.46	2.49E-09	4.84E-06	Fibro_Dup vs Fibro_Ctrl
17052060	7	136868652	137020255	*M2*	-2.65	3.33E-09	6.15E-06	Fibro_Dup vs Fibro_Ctrl
16751917	12	53985065	53990279	*HOXC10*	-1.43	3.52E-09	6.20E-06	Fibro_Dup vs Fibro_Ctrl
16983652	5	31193655	31329146	*CDH6*	-3.94	5.09E-09	8.54E-06	Fibro_Dup vs Fibro_Ctrl
17115271	X	153861514	153886173	*L1CAM*	-2.27	9.05E-09	1.34E-05	Fibro_Dup vs Fibro_Ctrl
16668220	1	109466478	109482142	*SYPL2*	-1.64	9.07E-09	1.34E-05	Fibro_Dup vs Fibro_Ctrl
17044177	7	22725884	22732002	*IL6*	-1.58	8.85E-09	1.34E-05	Fibro_Dup vs Fibro_Ctrl
16721884	11	9758292	9811319	*SBF2-AS1*	-1.67	1.32E-08	1.88E-05	Fibro_Dup vs Fibro_Ctrl
16759218	12	130953907	131141469	*ADGRD1*	-2.27	2.15E-08	2.94E-05	Fibro_Dup vs Fibro_Ctrl
17012761	6	133889121	133895553	*TCF21*	-1.92	2.41E-08	3.18E-05	Fibro_Dup vs Fibro_Ctrl
17012859	6	135851696	136195574	*PDE7B*	1.45	2.99E-08	3.81E-05	Fibro_Dup vs Fibro_Ctrl
16811085	15	68301704	68432163	*ITGA11*	3.04	3.14E-08	3.86E-05	Fibro_Dup vs Fibro_Ctrl
16821367	16	82035253	82098534	*HSD17B2*	-1.57	3.79E-08	4.38E-05	Fibro_Dup vs Fibro_Ctrl
16743721	11	102789910	102798235	*MMP1*	2.80	3.69E-08	4.38E-05	Fibro_Dup vs Fibro_Ctrl
17022980	6	116752197	116765723	*FAM162B*	-1.38	4.16E-08	4.65E-05	Fibro_Dup vs Fibro_Ctrl
16759262	12	130977562	130978768	*RP11-243M5.1*	-2.38	4.67E-08	5.04E-05	Fibro_Dup vs Fibro_Ctrl
16979339	4	119494395	119628991	*PDE5A*	-2.31	4.77E-08	5.04E-05	Fibro_Dup vs Fibro_Ctrl
16768738	12	95657807	95791152	*NTN4*	-2.21	5.72E-08	5.78E-05	Fibro_Dup vs Fibro_Ctrl
16975971	4	53459301	53701405	*LNX1*	1.24	5.79E-08	5.78E-05	Fibro_Dup vs Fibro_Ctrl
16711299	10	4200637	4243789	*LINC00702*	1.21	6.31E-08	6.13E-05	Fibro_Dup vs Fibro_Ctrl
17094889	9	72871728	72874109	*LINC01474*	-1.70	7.51E-08	7.11E-05	Fibro_Dup vs Fibro_Ctrl
16977986	4	89724099	89838315	*SNCA*	-1.46	7.82E-08	7.22E-05	Fibro_Dup vs Fibro_Ctrl
16922114	21	31873315	32044633	*HUNK*	2.00	8.53E-08	7.69E-05	Fibro_Dup vs Fibro_Ctrl
16958251	3	123610049	123884331	*MYLK*	-1.52	9.23E-08	8.12E-05	Fibro_Dup vs Fibro_Ctrl
17104313	X	67544032	67730876	*AR*	-1.54	1.18E-07	0.00010	Fibro_Dup vs Fibro_Ctrl
16919158	20	38128461	38166578	*TGM2*	-2.04	1.34E-07	0.00011	Fibro_Dup vs Fibro_Ctrl
16751910	12	53973126	53977643	*HOXC11*	-1.24	1.49E-07	0.00012	Fibro_Dup vs Fibro_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
17094498	9	40992261	41074625	*PGM5P2*	-1.59	1.53E-07	0.00012	Fibro_Dup vs Fibro_Ctrl
16762759	12	29500813	29784759	*TMTC1*	-2.52	1.90E-07	0.00015	Fibro_Dup vs Fibro_Ctrl
17051553	7	130293134	130324180	*CPA4*	-1.87	2.21E-07	0.00017	Fibro_Dup vs Fibro_Ctrl
16696120	1	168695457	168729264	*DPT*	3.22	2.25E-07	0.00017	Fibro_Dup vs Fibro_Ctrl
16948168	3	177683627	177691234	*RP11-114M1.1*	-1.32	2.40E-07	0.00018	Fibro_Dup vs Fibro_Ctrl
16972167	4	165361194	165498330	*CPE*	2.11	2.47E-07	0.00018	Fibro_Dup vs Fibro_Ctrl
16998682	5	107376889	107670895	*EFNA5*	1.69	2.96E-07	0.00021	Fibro_Dup vs Fibro_Ctrl
16815791	16	10386055	10483638	*ATF7IP2*	-1.89	3.04E-07	0.00021	Fibro_Dup vs Fibro_Ctrl
17085516	9	41073710	41076392	*PGM5-AS1*	-1.03	3.67E-07	0.00025	Fibro_Dup vs Fibro_Ctrl
16761221	12	9898673	9932381	*CLEC2A*	3.00	3.76E-07	0.00025	Fibro_Dup vs Fibro_Ctrl
17002667	5	169863715	169980740	*FAM196B*	1.36	3.83E-07	0.00025	Fibro_Dup vs Fibro_Ctrl
16761293	12	10158300	10172191	*OLR1*	-1.44	4.49E-07	0.00029	Fibro_Dup vs Fibro_Ctrl
17059451	7	87401697	87480435	*ABCB4*	1.22	4.45E-07	0.00029	Fibro_Dup vs Fibro_Ctrl
16986065	5	73120292	73131817	*TMEM171*	1.17	4.80E-07	0.00030	Fibro_Dup vs Fibro_Ctrl
16754269	12	72087266	72665642	*TRHDE*	-2.62	6.49E-07	0.00040	Fibro_Dup vs Fibro_Ctrl
17062127	7	117276631	117323289	*WNT2*	-1.55	6.71E-07	0.00040	Fibro_Dup vs Fibro_Ctrl
16736883	11	27696312	27877648	*RP11-587D21.4*	-1.36	6.69E-07	0.00040	Fibro_Dup vs Fibro_Ctrl
16697971	1	201896452	201946588	*LMOD1*	-1.48	7.57E-07	0.00044	Fibro_Dup vs Fibro_Ctrl
17023782	6	133435077	133511728	*TARID*	-0.86	7.88E-07	0.00045	Fibro_Dup vs Fibro_Ctrl
16668505	1	110082230	110110239	*LINC01397*	-1.58	8.32E-07	0.00047	Fibro_Dup vs Fibro_Ctrl
16836753	17	61452418	61485108	*TBX4*	-1.01	8.25E-07	0.00047	Fibro_Dup vs Fibro_Ctrl
16894824	2	19351485	19358653	*OSR1*	1.73	1.05E-06	0.00058	Fibro_Dup vs Fibro_Ctrl
16667516	1	99263953	99309590	*PLPPR4*	-2.26	1.10E-06	0.00059	Fibro_Dup vs Fibro_Ctrl
16767606	12	72253507	72274907	*TRHDE-AS1*	-1.96	1.11E-06	0.00059	Fibro_Dup vs Fibro_Ctrl
16787831	14	94146128	94279735	*PPP4R4*	-1.81	1.12E-06	0.00059	Fibro_Dup vs Fibro_Ctrl
16970118	4	118888841	119061247	*SYNPO2*	-1.88	1.24E-06	0.00065	Fibro_Dup vs Fibro_Ctrl
17045897	7	47795291	47819847	*C7orf69*	-1.67	1.35E-06	0.00069	Fibro_Dup vs Fibro_Ctrl
16965142	4	14111968	14140052	*LINC01085*	-1.28	1.55E-06	0.00079	Fibro_Dup vs Fibro_Ctrl
16981706	4	176213673	176277571	*ASB5*	-1.07	1.59E-06	0.00080	Fibro_Dup vs Fibro_Ctrl
16778061	13	35473789	35476695	*MAB21L1*	-2.78	1.62E-06	0.00080	Fibro_Dup vs Fibro_Ctrl
16887857	2	174326027	174330643	*LINC01305*	1.30	1.76E-06	0.00086	Fibro_Dup vs Fibro_Ctrl
16687737	1	56494747	56645301	*PLPP3*	1.41	1.90E-06	0.00091	Fibro_Dup vs Fibro_Ctrl
16964888	4	8269762	8307111	*HTRA3*	1.20	2.00E-06	0.00095	Fibro_Dup vs Fibro_Ctrl
16978354	4	100395341	100880126	*EMCN*	0.99	2.05E-06	0.00096	Fibro_Dup vs Fibro_Ctrl
16986138	5	73626158	73942168	*ARHGEF28*	-0.91	2.10E-06	0.00097	Fibro_Dup vs Fibro_Ctrl
16961331	3	169083499	169663775	*MECOM*	-1.60	2.32E-06	0.00106	Fibro_Dup vs Fibro_Ctrl
16703036	10	19816239	20280357	*PLXDC2*	2.73	2.74E-06	0.00123	Fibro_Dup vs Fibro_Ctrl
16732394	11	120213736	120217513	*OAF*	-1.02	2.99E-06	0.00133	Fibro_Dup vs Fibro_Ctrl
16852858	18	63752935	63805370	*SERPINB7*	-1.82	3.05E-06	0.00134	Fibro_Dup vs Fibro_Ctrl
16701037	1	240489573	240612162	*GREM2*	1.87	3.30E-06	0.00143	Fibro_Dup vs Fibro_Ctrl
17023799	6	133987581	134052651	*SLC2A12*	-2.01	3.35E-06	0.00144	Fibro_Dup vs Fibro_Ctrl
16852676	18	59220168	59230774	*GRP*	1.04	3.44E-06	0.00146	Fibro_Dup vs Fibro_Ctrl
17078254	8	73064540	73124088	*SBSPON*	-1.01	3.67E-06	0.00152	Fibro_Dup vs Fibro_Ctrl
16689354	1	89107627	89126116	*GBP2*	1.21	3.63E-06	0.00152	Fibro_Dup vs Fibro_Ctrl
16935905	22	44243667	44312851	*KIAA1644*	1.81	4.54E-06	0.00186	Fibro_Dup vs Fibro_Ctrl
16946707	3	148697784	148743164	*AGTR1*	0.90	5.05E-06	0.00205	Fibro_Dup vs Fibro_Ctrl
17068349	8	40153468	40155308	*C8orf4*	-2.16	5.72E-06	0.00229	Fibro_Dup vs Fibro_Ctrl
16712168	10	16823964	17129831	*CUBN*	1.36	7.88E-06	0.00313	Fibro_Dup vs Fibro_Ctrl
16731297	11	112961247	113278436	*NCAM1*	1.52	8.72E-06	0.00343	Fibro_Dup vs Fibro_Ctrl
16769552	12	105304867	105327017	*KCCAT198*	1.65	9.25E-06	0.00360	Fibro_Dup vs Fibro_Ctrl
17112235	X	75368427	75523502	*ZDHHC15*	1.52	9.61E-06	0.00370	Fibro_Dup vs Fibro_Ctrl
16904667	2	166195185	166375993	*SCN9A*	-1.90	1.00E-05	0.00381	Fibro_Dup vs Fibro_Ctrl
16802653	15	71096952	71783383	*THSD4*	-1.93	1.02E-05	0.00381	Fibro_Dup vs Fibro_Ctrl
16769761	12	108589846	108598320	*TMEM119*	1.05	1.01E-05	0.00381	Fibro_Dup vs Fibro_Ctrl
16805192	15	92393828	92472410	*ST8SIA2*	1.22	1.15E-05	0.00424	Fibro_Dup vs Fibro_Ctrl
16922759	21	38157034	38301826	*KCNJ15*	1.00	1.19E-05	0.00432	Fibro_Dup vs Fibro_Ctrl
16984294	5	40841184	40860173	*CARD6*	1.00	1.19E-05	0.00432	Fibro_Dup vs Fibro_Ctrl
16897834	2	55865962	55924163	*EFEMP1*	-2.34	1.25E-05	0.00450	Fibro_Dup vs Fibro_Ctrl
16987531	5	96875939	96919716	*ERAP2*	1.30	1.30E-05	0.00461	Fibro_Dup vs Fibro_Ctrl
16967843	4	74365143	74388760	*EREG*	-1.30	1.42E-05	0.00500	Fibro_Dup vs Fibro_Ctrl
17004198	6	1389834	1395597	*FOXF2*	-1.56	1.49E-05	0.00519	Fibro_Dup vs Fibro_Ctrl
16994597	5	16473038	16617058	*FAM134B*	1.56	1.51E-05	0.00522	Fibro_Dup vs Fibro_Ctrl
16998293	5	95751319	95823005	*GLRX*	-0.93	1.56E-05	0.00534	Fibro_Dup vs Fibro_Ctrl
16986098	5	73446357	73446984	*FOXD1-AS1*	1.33	1.59E-05	0.00537	Fibro_Dup vs Fibro_Ctrl
16972446	4	173527270	173591324	*HAND2-AS1*	1.37	1.77E-05	0.00594	Fibro_Dup vs Fibro_Ctrl
16872884	19	42902079	42918010	*PSG6*	-1.56	1.82E-05	0.00605	Fibro_Dup vs Fibro_Ctrl
17063221	7	135728348	135748846	*FAM180A*	1.02	1.94E-05	0.00638	Fibro_Dup vs Fibro_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16989445	5	135265437	135267895	*C5orf66*	-0.65	2.05E-05	0.00670	Fibro_Dup vs Fibro_Ctrl
16852982	18	69401048	69849087	*DOK6*	1.23	2.11E-05	0.00684	Fibro_Dup vs Fibro_Ctrl
16919915	20	47209214	47356823	*ZMYND8*	-0.70	2.24E-05	0.00718	Fibro_Dup vs Fibro_Ctrl
17077376	8	56494282	56559823	*LINC00968*	-1.36	2.31E-05	0.00736	Fibro_Dup vs Fibro_Ctrl
16908138	2	215942591	216034096	*MREG*	1.11	2.34E-05	0.00740	Fibro_Dup vs Fibro_Ctrl
16921627	21	16070522	16609775	*MIR99AHG*	-1.30	2.63E-05	0.00823	Fibro_Dup vs Fibro_Ctrl
16962536	3	187216085	187292022	*MASP1*	-1.19	2.74E-05	0.00850	Fibro_Dup vs Fibro_Ctrl
16948161	3	177441921	177752704	*LINC00578*	-0.94	2.78E-05	0.00856	Fibro_Dup vs Fibro_Ctrl
16802251	15	67063763	67195195	*SMAD3*	-1.00	2.98E-05	0.00908	Fibro_Dup vs Fibro_Ctrl
16677451	1	215005542	215237093	*KCNK2*	1.34	3.16E-05	0.00956	Fibro_Dup vs Fibro_Ctrl
16872845	19	42752686	42765696	*PSG8*	-1.09	3.24E-05	0.00971	Fibro_Dup vs Fibro_Ctrl
16924602	21	26836287	26845409	*ADAMTS1*	-1.77	3.37E-05	0.00995	Fibro_Dup vs Fibro_Ctrl
16821377	16	82626794	83796610	*CDH13*	1.90	3.39E-05	0.00995	Fibro_Dup vs Fibro_Ctrl
16791319	14	24318349	24335093	*ADCY4*	-0.71	3.38E-05	0.00995	Fibro_Dup vs Fibro_Ctrl
16753158	12	57726240	57728356	*AGAP2-AS1*	1.11	3.55E-05	0.01032	Fibro_Dup vs Fibro_Ctrl
17115375	X	153935263	153944780	*RENBP*	0.86	3.63E-05	0.01047	Fibro_Dup vs Fibro_Ctrl
16664828	1	53062052	53085502	*PODN*	-1.05	3.70E-05	0.01053	Fibro_Dup vs Fibro_Ctrl
16945768	3	132597237	132618967	*ACKR4*	1.83	3.71E-05	0.01053	Fibro_Dup vs Fibro_Ctrl
16852871	18	63871692	63903890	*SERPINB2*	-2.04	3.85E-05	0.01084	Fibro_Dup vs Fibro_Ctrl
16855673	18	63123346	63320128	*BCL2*	1.28	3.94E-05	0.01102	Fibro_Dup vs Fibro_Ctrl
16877616	2	23385217	23708613	*KLHL29*	-0.96	4.34E-05	0.01204	Fibro_Dup vs Fibro_Ctrl
17070013	8	73294602	73325285	*RDH10*	-1.08	4.40E-05	0.01213	Fibro_Dup vs Fibro_Ctrl
17059828	7	93885397	93890991	*TFPI2*	-2.22	4.56E-05	0.01249	Fibro_Dup vs Fibro_Ctrl
16905436	2	176629581	176637931	*LINC01116*	1.26	4.77E-05	0.01289	Fibro_Dup vs Fibro_Ctrl
16848219	17	69078706	69141874	*ABCA6*	1.60	4.78E-05	0.01289	Fibro_Dup vs Fibro_Ctrl
17097130	9	110539690	110539798	*RNU6-124P*	-1.47	4.97E-05	0.01330	Fibro_Dup vs Fibro_Ctrl
16775491	13	75871038	75883812	*LMO7DN*	-1.09	5.05E-05	0.01343	Fibro_Dup vs Fibro_Ctrl
16843241	17	32492610	32876884	*MYO1D*	1.53	5.49E-05	0.01448	Fibro_Dup vs Fibro_Ctrl
16980160	4	142023160	142847432	*INPP4B*	-1.16	5.61E-05	0.01468	Fibro_Dup vs Fibro_Ctrl
16802960	15	74173671	74176872	*ISLR*	0.99	5.85E-05	0.01521	Fibro_Dup vs Fibro_Ctrl
16718622	10	114431110	114684655	*ABLIM1*	-0.89	5.97E-05	0.01542	Fibro_Dup vs Fibro_Ctrl
16920156	20	49503874	49568170	*PTGIS*	-1.61	6.12E-05	0.01569	Fibro_Dup vs Fibro_Ctrl
16964950	4	8592660	8619761	*CPZ*	-1.17	6.29E-05	0.01602	Fibro_Dup vs Fibro_Ctrl
16706117	10	73772291	73779092	*FUT11*	0.74	6.63E-05	0.01678	Fibro_Dup vs Fibro_Ctrl
16978236	4	99304964	99321415	*ADH1B*	-1.19	6.71E-05	0.01687	Fibro_Dup vs Fibro_Ctrl
16837191	17	67377808	67693531	*PITPNC1*	1.00	6.88E-05	0.01714	Fibro_Dup vs Fibro_Ctrl
17019698	6	46220730	46492067	*RCAN2*	1.91	6.92E-05	0.01714	Fibro_Dup vs Fibro_Ctrl
16960911	3	158645822	158672693	*LXN*	-1.44	7.05E-05	0.01735	Fibro_Dup vs Fibro_Ctrl
16906419	2	189560579	189583758	*SLC40A1*	-1.02	7.10E-05	0.01736	Fibro_Dup vs Fibro_Ctrl
16705011	10	52314281	52318042	*DKK1*	-1.18	7.16E-05	0.01740	Fibro_Dup vs Fibro_Ctrl
16712086	10	15513949	15720125	*ITGA8*	2.29	7.32E-05	0.01767	Fibro_Dup vs Fibro_Ctrl
16666357	1	77281977	77559969	*AK5*	1.20	7.50E-05	0.01791	Fibro_Dup vs Fibro_Ctrl
17076244	8	37516450	37517021	*LINC01605*	0.96	7.52E-05	0.01791	Fibro_Dup vs Fibro_Ctrl
16937741	3	12287368	12434356	*PPARG*	-1.08	8.15E-05	0.01930	Fibro_Dup vs Fibro_Ctrl
16999557	5	128371099	128371212	*Y_RNA*	-1.98	8.51E-05	0.02001	Fibro_Dup vs Fibro_Ctrl
16876777	2	6917392	7068286	*RNF144A*	1.09	8.63E-05	0.02017	Fibro_Dup vs Fibro_Ctrl
16792211	14	38254103	38256369	*CLEC14A*	1.01	8.74E-05	0.02031	Fibro_Dup vs Fibro_Ctrl
16706818	10	85537379	85537483	*RNU6-325P*	-1.17	8.91E-05	0.02056	Fibro_Dup vs Fibro_Ctrl
16707503	10	92831178	93059494	*EXOC6*	-1.13	9.17E-05	0.02104	Fibro_Dup vs Fibro_Ctrl
16671065	1	152776372	152776969	*LCE1F*	-1.09	0.00010	0.02371	Fibro_Dup vs Fibro_Ctrl
16775498	13	75878713	75881127	*LMO7DN-IT1*	-0.91	0.00011	0.02452	Fibro_Dup vs Fibro_Ctrl
17072920	8	133191039	133231690	*WISP1*	1.14	0.00011	0.02456	Fibro_Dup vs Fibro_Ctrl
16829728	17	3474110	3503419	*ASPA*	1.30	0.00011	0.02566	Fibro_Dup vs Fibro_Ctrl
16872926	19	43166256	43186536	*PSG5*	-1.81	0.00012	0.02603	Fibro_Dup vs Fibro_Ctrl
16983848	5	34656237	34832627	*RAI14*	-0.86	0.00012	0.02658	Fibro_Dup vs Fibro_Ctrl
16923705	21	45073853	45226563	*ADARB1*	0.79	0.00012	0.02672	Fibro_Dup vs Fibro_Ctrl
16874097	19	48813017	48836677	*HSD17B14*	1.62	0.00013	0.02747	Fibro_Dup vs Fibro_Ctrl
16665254	1	59020387	59044614	*LINC01358*	0.83	0.00013	0.02822	Fibro_Dup vs Fibro_Ctrl
16899764	2	85839144	85889034	*ST3GAL5*	1.29	0.00013	0.02838	Fibro_Dup vs Fibro_Ctrl
16755712	12	100717526	101128641	*ANO4*	-1.58	0.00013	0.02841	Fibro_Dup vs Fibro_Ctrl
16957675	3	115802363	117997592	*LSAMP*	1.34	0.00013	0.02851	Fibro_Dup vs Fibro_Ctrl
17072375	8	123416725	123467230	*WDYHV1*	0.73	0.00014	0.02990	Fibro_Dup vs Fibro_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	0.70	0.00014	0.02991	Fibro_Dup vs Fibro_Ctrl
16666336	1	76867441	77065711	*ST6GALNAC5*	1.38	0.00015	0.03110	Fibro_Dup vs Fibro_Ctrl
16779839	13	76880169	76886405	*KCTD12*	1.63	0.00015	0.03135	Fibro_Dup vs Fibro_Ctrl
16907912	2	212999684	213152427	*IKZF2*	-0.69	0.00015	0.03135	Fibro_Dup vs Fibro_Ctrl
16906749	2	196199253	196593692	*HECW2*	0.93	0.00016	0.03221	Fibro_Dup vs Fibro_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16921827	21	25561909	25575168	*MIR155HG*	0.69	0.00016	0.03310	Fibro_Dup vs Fibro_Ctrl
16902199	2	118842171	118848183	*EN1*	1.17	0.00016	0.03330	Fibro_Dup vs Fibro_Ctrl
16675197	1	186828900	186988981	*PLA2G4A*	2.06	0.00016	0.03330	Fibro_Dup vs Fibro_Ctrl
16666416	1	77888513	77943895	*NEXN*	-1.50	0.00017	0.03449	Fibro_Dup vs Fibro_Ctrl
16756163	12	104215779	104350284	*TXNRD1*	-0.78	0.00017	0.03449	Fibro_Dup vs Fibro_Ctrl
16965252	4	15702950	15738313	*BST1*	0.72	0.00017	0.03449	Fibro_Dup vs Fibro_Ctrl
17115782	X	155276211	155334677	*CLIC2*	1.48	0.00018	0.03553	Fibro_Dup vs Fibro_Ctrl
16844775	17	41523617	41528389	*KRT19*	-1.27	0.00018	0.03602	Fibro_Dup vs Fibro_Ctrl
17107804	X	151695124	151705924	*PRRG3*	0.87	0.00018	0.03602	Fibro_Dup vs Fibro_Ctrl
16817254	16	27313926	27364778	*IL4R*	0.82	0.00019	0.03631	Fibro_Dup vs Fibro_Ctrl
17047112	7	74289473	74405943	*CLIP2*	0.85	0.00019	0.03715	Fibro_Dup vs Fibro_Ctrl
17095499	9	86944362	86947189	*GAS1*	1.30	0.00020	0.03845	Fibro_Dup vs Fibro_Ctrl
16713052	10	31693084	31707388	*LOC101929951*	1.07	0.00020	0.03855	Fibro_Dup vs Fibro_Ctrl
17022018	6	100385015	100464929	*SIM1*	-1.75	0.00020	0.03892	Fibro_Dup vs Fibro_Ctrl
17014091	6	157823249	157945077	*SNX9*	0.89	0.00020	0.03892	Fibro_Dup vs Fibro_Ctrl
17012047	6	117453817	117569858	*DCBLD1*	-0.97	0.00021	0.03895	Fibro_Dup vs Fibro_Ctrl
16995177	5	35048743	35230721	*PRLR*	1.65	0.00021	0.03948	Fibro_Dup vs Fibro_Ctrl
16907044	2	199269500	199471266	*SATB2*	1.01	0.00022	0.04093	Fibro_Dup vs Fibro_Ctrl
16906991	2	197570800	197676045	*RFTN2*	1.77	0.00022	0.04130	Fibro_Dup vs Fibro_Ctrl
16765464	12	53983450	53985519	*HOXC-AS3*	-0.90	0.00022	0.04130	Fibro_Dup vs Fibro_Ctrl
17061881	7	111726108	112206411	*DOCK4*	1.60	0.00023	0.04157	Fibro_Dup vs Fibro_Ctrl
16769770	12	108621895	108633959	*SELPLG*	-0.74	0.00023	0.04191	Fibro_Dup vs Fibro_Ctrl
16677259	1	212432887	212446379	*NENF*	0.84	0.00023	0.04227	Fibro_Dup vs Fibro_Ctrl
17082976	9	72701	88826	*PGM5P3-AS1*	-0.56	0.00024	0.04310	Fibro_Dup vs Fibro_Ctrl
17011784	6	112236596	112306683	*LOC101927640*	0.59	0.00024	0.04418	Fibro_Dup vs Fibro_Ctrl
17005276	6	21593741	21598619	*SOX4*	1.55	0.00025	0.04425	Fibro_Dup vs Fibro_Ctrl
16777794	13	29764371	29850684	*UBL3*	1.31	0.00026	0.04572	Fibro_Dup vs Fibro_Ctrl
16706610	10	80207710	80219657	*LINC00857*	-0.71	0.00026	0.04572	Fibro_Dup vs Fibro_Ctrl
17113463	X	115110616	115234072	*LRCH2*	-1.36	0.00026	0.04594	Fibro_Dup vs Fibro_Ctrl
16795819	14	91232532	91253925	*GPR68*	0.77	0.00026	0.04594	Fibro_Dup vs Fibro_Ctrl
16691576	1	118883043	118989556	*TBX15*	1.17	0.00026	0.04598	Fibro_Dup vs Fibro_Ctrl
16980762	4	153780590	153789120	*SFRP2*	2.08	0.00027	0.04681	Fibro_Dup vs Fibro_Ctrl
17078539	8	80628451	80874781	*ZNF704*	1.34	0.00027	0.04681	Fibro_Dup vs Fibro_Ctrl
16754646	12	80405994	80680234	*PTPRQ*	-1.19	0.00027	0.04720	Fibro_Dup vs Fibro_Ctrl
17096493	9	98288082	98709197	*GABBR2*	-1.43	0.00027	0.04721	Fibro_Dup vs Fibro_Ctrl
17067566	8	30384428	30572261	*RBPMS*	-1.32	0.00028	0.04805	Fibro_Dup vs Fibro_Ctrl
16972249	4	168497066	168928457	*PALLD*	-0.87	0.00029	0.04880	Fibro_Dup vs Fibro_Ctrl
17021596	6	89364616	89412276	*RRAGD*	1.40	0.00029	0.04933	Fibro_Dup vs Fibro_Ctrl
16690211	1	102876467	103108496	*COL11A1*	-2.00	0.00029	0.04981	Fibro_Dup vs Fibro_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	1.24	5.69E-12	2.10E-07	Fibro_Dup vs Fibro_Del
16802429	15	68818221	69062743	*NOX5*	0.95	2.44E-09	4.51E-05	Fibro_Dup vs Fibro_Del
17058484	7	73536353	73558002	*BCL7B*	1.02	3.42E-08	0.00042	Fibro_Dup vs Fibro_Del
17047073	7	74174245	74197101	*EIF4H*	1.09	5.48E-08	0.00051	Fibro_Dup vs Fibro_Del
16872871	19	42866461	42879719	*PSG1*	-1.94	1.79E-07	0.00104	Fibro_Dup vs Fibro_Del
17047112	7	74289473	74405943	*CLIP2*	1.05	1.58E-07	0.00104	Fibro_Dup vs Fibro_Del
17047045	7	74082933	74122525	*LIMK1*	1.05	1.97E-07	0.00104	Fibro_Dup vs Fibro_Del
16920156	20	49503874	49568170	*PTGIS*	-1.71	6.35E-07	0.00213	Fibro_Dup vs Fibro_Del
16949827	3	195147871	195152790	*XXYLT1-AS2*	-1.24	4.76E-07	0.00213	Fibro_Dup vs Fibro_Del
16852676	18	59220168	59230774	*GRP*	0.93	6.25E-07	0.00213	Fibro_Dup vs Fibro_Del
17073220	8	142700111	142705127	*LY6K*	0.95	6.17E-07	0.00213	Fibro_Dup vs Fibro_Del
17058499	7	73568944	73578791	*TBL2*	1.06	1.30E-06	0.00401	Fibro_Dup vs Fibro_Del
17006911	6	31897785	31945672	*C2*	-1.09	1.45E-06	0.00410	Fibro_Dup vs Fibro_Del
17067748	8	32996178	33044855	*LOC105379362*	1.32	1.55E-06	0.00410	Fibro_Dup vs Fibro_Del
16910000	2	238788648	238789716	*AC144525.1*	-0.74	3.35E-06	0.00825	Fibro_Dup vs Fibro_Del
16791371	14	24336021	24340045	*RIPK3*	-0.87	3.80E-06	0.00878	Fibro_Dup vs Fibro_Del
16886417	2	149038107	149215262	*LYPD6B*	0.91	4.27E-06	0.00928	Fibro_Dup vs Fibro_Del
16791319	14	24318349	24335093	*ADCY4*	-0.65	5.20E-06	0.01067	Fibro_Dup vs Fibro_Del
16689330	1	89020246	89020365	*GBP3*	-0.98	6.26E-06	0.01217	Fibro_Dup vs Fibro_Del
17047138	7	74453790	74602605	*GTF2IRD1*	0.60	9.15E-06	0.01689	Fibro_Dup vs Fibro_Del
16870453	19	18782773	18791314	*COMP*	-0.95	9.71E-06	0.01707	Fibro_Dup vs Fibro_Del
16978354	4	100395341	100880126	*EMCN*	-0.73	1.14E-05	0.01910	Fibro_Dup vs Fibro_Del
16842147	17	18528586	18625617	*CCDC144B*	1.55	1.63E-05	0.02611	Fibro_Dup vs Fibro_Del
16924101	21	9068356	9129760	*TEKT4P2*	1.51	1.90E-05	0.02923	Fibro_Dup vs Fibro_Del
16674292	1	178725165	179096418	*RALGPS2*	-0.74	2.18E-05	0.03218	Fibro_Dup vs Fibro_Del
16835158	17	47253842	47312711	*ITGB3*	-1.20	2.64E-05	0.03747	Fibro_Dup vs Fibro_Del
16997383	5	76615482	76623434	*F2RL2*	1.68	3.03E-05	0.04150	Fibro_Dup vs Fibro_Del
16664828	1	53062052	53085502	*PODN*	-0.87	3.17E-05	0.04185	Fibro_Dup vs Fibro_Del

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
17012664	6	131901963	131920565	*RP1-131F15.3*	2.29	1.12E-10	4.15E-06	iPS_Del vs iPS_Ctrl
17011157	6	97816597	97998912	*LOC101927335*	1.78	4.72E-08	0.00087	iPS_Del vs iPS_Ctrl
17095968	9	93318199	93346414	*C9orf129*	1.50	1.16E-07	0.00143	iPS_Del vs iPS_Ctrl
16728141	11	68683779	68691175	*GAL*	1.33	3.77E-07	0.00348	iPS_Del vs iPS_Ctrl
16686271	1	44721786	44721902	*RNU5F-1*	-3.21	3.10E-06	0.01432	iPS_Del vs iPS_Ctrl
17023327	6	125578567	125749186	*LOC105377986*	1.05	2.06E-06	0.01432	iPS_Del vs iPS_Ctrl
17055657	7	19918981	20217404	*MACC1*	1.36	3.09E-06	0.01432	iPS_Del vs iPS_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	-0.71	2.86E-06	0.01432	iPS_Del vs iPS_Ctrl
16774623	13	46553161	46753041	*LRCH1*	1.01	4.44E-06	0.01820	iPS_Del vs iPS_Ctrl
16966304	4	41359607	41700044	*LIMCH1*	-1.40	5.06E-06	0.01868	iPS_Del vs iPS_Ctrl
16855469	18	57054572	57072119	*LINC-ROR*	1.49	5.59E-06	0.01875	iPS_Del vs iPS_Ctrl
16933429	22	27675822	27677344	*RP1-213J1P_B.1*	1.44	7.16E-06	0.02203	iPS_Del vs iPS_Ctrl
16754565	12	79540203	79550535	*RP11-359M6.1*	2.48	8.84E-06	0.02411	iPS_Del vs iPS_Ctrl
16725084	11	59443144	59444194	*OR5A1*	-0.63	9.14E-06	0.02411	iPS_Del vs iPS_Ctrl
16791909	14	34752731	34875647	*BAZ1A*	0.91	1.16E-05	0.02847	iPS_Del vs iPS_Ctrl
16870854	19	22179101	22196951	*ZNF676*	-3.63	1.26E-05	0.02909	iPS_Del vs iPS_Ctrl
16870815	19	21569903	21587322	*RP11-678G14.3*	-0.85	1.67E-05	0.03623	iPS_Del vs iPS_Ctrl
16666543	1	81251789	81252096	*RN7SKP247*	1.20	1.99E-05	0.03867	iPS_Del vs iPS_Ctrl
17051217	7	128709061	128731743	*FAM71F1*	0.81	1.93E-05	0.03867	iPS_Del vs iPS_Ctrl
17047138	7	74453790	74602605	*GTF2IRD1*	-0.55	2.34E-05	0.04326	iPS_Del vs iPS_Ctrl
16670997	1	152188851	152366836	*FLG-AS1*	2.06	9.61E-09	0.00035	iPS_Dup vs iPS_Ctrl
16865945	19	56595296	56624182	*ZNF71*	0.90	1.10E-07	0.00202	iPS_Dup vs iPS_Ctrl
17017900	6	32578769	32589848	*HLA-DRB1*	-2.07	1.39E-06	0.01586	iPS_Dup vs iPS_Ctrl
16671013	1	152205858	152207057	*LOC105371443*	1.59	1.72E-06	0.01586	iPS_Dup vs iPS_Ctrl
16993311	5	178941191	178966433	*ZNF454*	1.26	5.99E-06	0.03886	iPS_Dup vs iPS_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	0.69	6.31E-06	0.03886	iPS_Dup vs iPS_Ctrl
17051217	7	128709061	128731743	*FAM71F1*	0.87	8.44E-06	0.04451	iPS_Dup vs iPS_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	1.40	2.66E-19	9.81E-15	iPS_Dup vs iPS_Del
17047073	7	74174245	74197101	*EIF4H*	1.33	4.98E-15	9.20E-11	iPS_Dup vs iPS_Del
17012664	6	131901963	131920565	*RP1-131F15.3*	-2.07	1.66E-13	1.65E-09	iPS_Dup vs iPS_Del
17047138	7	74453790	74602605	*GTF2IRD1*	0.86	1.78E-13	1.65E-09	iPS_Dup vs iPS_Del
17047045	7	74082933	74122525	*LIMK1*	1.13	2.92E-12	2.16E-08	iPS_Dup vs iPS_Del
17058499	7	73568944	73578791	*TBL2*	1.19	1.60E-11	9.84E-08	iPS_Dup vs iPS_Del
17058582	7	73736094	73738867	*ABHD11*	0.96	3.07E-11	1.62E-07	iPS_Dup vs iPS_Del
17058484	7	73536353	73558002	*BCL7B*	0.92	6.12E-11	2.83E-07	iPS_Dup vs iPS_Del
16670997	1	152188851	152366836	*FLG-AS1*	1.68	4.91E-10	2.02E-06	iPS_Dup vs iPS_Del
17058645	7	74689057	74728849	*LOC101926943*	0.78	1.75E-09	6.45E-06	iPS_Dup vs iPS_Del
17072014	8	114282067	114295839	*RP11-267L5.1*	-2.53	1.24E-08	3.98E-05	iPS_Dup vs iPS_Del
17047086	7	74191198	74191294	*MIR590*	2.07	1.29E-08	3.98E-05	iPS_Dup vs iPS_Del
16666543	1	81251789	81252096	*RN7SKP247*	-1.25	1.66E-08	4.73E-05	iPS_Dup vs iPS_Del
17051046	7	125917871	125933832	*AC005276.1*	-2.92	2.64E-08	6.98E-05	iPS_Dup vs iPS_Del
16865110	19	53788705	53788773	*MIR373*	-1.27	7.48E-08	0.00018	iPS_Dup vs iPS_Del
17058617	7	74231499	74254444	*RFC2*	1.26	1.67E-07	0.00038	iPS_Dup vs iPS_Del
16938309	3	21542816	21579959	*ZNF385D-AS1*	-2.39	1.90E-07	0.00041	iPS_Dup vs iPS_Del
17046948	7	73683025	73705163	*WBSCR22*	0.39	2.11E-07	0.00043	iPS_Dup vs iPS_Del
16686271	1	44721786	44721902	*RNU5F-1*	2.59	5.43E-07	0.00095	iPS_Dup vs iPS_Del
16869060	19	12043805	12052967	*ZNF878*	-0.82	5.13E-07	0.00095	iPS_Dup vs iPS_Del
17047112	7	74289473	74405943	*CLIP2*	0.70	5.40E-07	0.00095	iPS_Dup vs iPS_Del
17095968	9	93318199	93346414	*C9orf129*	-1.02	6.06E-07	0.00097	iPS_Dup vs iPS_Del
16671013	1	152205858	152207057	*LOC105371443*	1.22	6.05E-07	0.00097	iPS_Dup vs iPS_Del
16874514	19	50152548	50163281	*IZUMO2*	-0.68	8.53E-07	0.00131	iPS_Dup vs iPS_Del
16758253	12	122063306	122068764	*LOC100506691*	0.77	1.15E-06	0.00170	iPS_Dup vs iPS_Del
16875219	19	53793584	53824394	*NLRP12*	-0.75	1.32E-06	0.00188	iPS_Dup vs iPS_Del
16990478	5	142745600	142760998	*LOC101926975*	-1.09	1.55E-06	0.00213	iPS_Dup vs iPS_Del
17058427	7	73302517	73308867	*NSUN5*	0.88	2.49E-06	0.00328	iPS_Dup vs iPS_Del
17046982	7	73799542	73832687	*CLDN4*	0.67	4.14E-06	0.00527	iPS_Dup vs iPS_Del
17047176	7	74657667	74760696	*GTF2I*	0.95	6.26E-06	0.00771	iPS_Dup vs iPS_Del
16989557	5	136466677	136520298	*RP11-114H21.2*	1.47	7.68E-06	0.00886	iPS_Dup vs iPS_Del
17066573	8	22275297	22357563	*PIWIL2*	-0.96	7.49E-06	0.00886	iPS_Dup vs iPS_Del
17062536	7	125229579	125264291	*LOC101928254*	-1.45	8.73E-06	0.00977	iPS_Dup vs iPS_Del
16819789	16	65371100	65393701	*RP11-256I9.3*	-1.14	1.24E-05	0.01351	iPS_Dup vs iPS_Del
16990267	5	141135218	141138122	*PCDHB5*	1.83	1.44E-05	0.01516	iPS_Dup vs iPS_Del
17015788	6	14280127	14285219	*LINC01108*	-2.09	1.56E-05	0.01553	iPS_Dup vs iPS_Del
16855469	18	57054572	57072119	*LINC-ROR*	-1.03	1.63E-05	0.01553	iPS_Dup vs iPS_Del
16933749	22	30008742	30080480	*HORMAD2-AS1*	-0.57	1.57E-05	0.01553	iPS_Dup vs iPS_Del
17058524	7	73593194	73624543	*MLXIPL*	0.69	1.64E-05	0.01553	iPS_Dup vs iPS_Del
17058557	7	73699205	73719687	*STX1A*	0.44	1.94E-05	0.01788	iPS_Dup vs iPS_Del

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16792300	14	41587861	41604856	*LOC105370469*	-1.32	2.36E-05	0.02127	iPS_Dup vs iPS_Del
16865078	19	53731006	53731090	*MIR518A1*	-0.74	3.45E-05	0.03030	iPS_Dup vs iPS_Del
16722987	11	24496970	25082640	*LUZP2*	1.31	4.04E-05	0.03469	iPS_Dup vs iPS_Del
16783739	14	41607561	41904549	*LRFN5*	-1.01	4.52E-05	0.03709	iPS_Dup vs iPS_Del
16860322	19	23914876	23945159	*ZNF726*	0.89	4.42E-05	0.03709	iPS_Dup vs iPS_Del
16870838	19	21966095	22010943	*ZNF208*	2.79	4.65E-05	0.03734	iPS_Dup vs iPS_Del
16981099	4	163323961	163344832	*NPY1R*	0.75	4.84E-05	0.03802	iPS_Dup vs iPS_Del
16855049	18	47028202	47030078	*TCEB3C*	-0.92	5.03E-05	0.03872	iPS_Dup vs iPS_Del
17011157	6	97816597	97998912	*LOC101927335*	-0.90	5.48E-05	0.04130	iPS_Dup vs iPS_Del
17058551	7	73680918	73683453	*DNAJC30*	0.65	5.63E-05	0.04155	iPS_Dup vs iPS_Del
16890027	2	206866795	206868289	*LOC105373850*	-1.25	6.34E-05	0.04593	iPS_Dup vs iPS_Del
17058460	7	73440398	73522285	*BAZ1B*	-1.03	5.62E-10	2.07E-05	NPC_Del vs NPC_Ctrl
17058484	7	73536353	73558002	*BCL7B*	-0.85	7.55E-07	0.01395	NPC_Del vs NPC_Ctrl
17047138	7	74453790	74602605	*GTF2IRD1*	-0.65	1.14E-06	0.01405	NPC_Del vs NPC_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	1.28	9.65E-18	3.56E-13	NPC_Dup vs NPC_Del
17047073	7	74174245	74197101	*EIF4H*	1.27	3.02E-14	3.72E-10	NPC_Dup vs NPC_Del
17047138	7	74453790	74602605	*GTF2IRD1*	0.91	2.11E-14	3.72E-10	NPC_Dup vs NPC_Del
17058484	7	73536353	73558002	*BCL7B*	1.12	1.16E-13	1.07E-09	NPC_Dup vs NPC_Del
17046948	7	73683025	73705163	*WBSCR22*	0.56	1.60E-11	1.18E-07	NPC_Dup vs NPC_Del
16858546	19	11848761	11869491	*ZNF439*	-1.37	7.68E-11	4.73E-07	NPC_Dup vs NPC_Del
17047045	7	74082933	74122525	*LIMK1*	0.95	5.50E-10	2.90E-06	NPC_Dup vs NPC_Del
17058499	7	73568944	73578791	*TBL2*	1.04	7.61E-10	3.51E-06	NPC_Dup vs NPC_Del
16959507	3	135261028	135439888	*LOC105374122*	1.04	1.70E-09	6.99E-06	NPC_Dup vs NPC_Del
16713330	10	37775206	37784131	*ZNF248*	1.46	1.33E-08	4.90E-05	NPC_Dup vs NPC_Del
16854792	18	39206924	39800318	*MIR924HG*	-2.21	1.53E-08	5.13E-05	NPC_Dup vs NPC_Del
17047086	7	74191198	74191294	*MIR590*	1.98	3.80E-08	0.00012	NPC_Dup vs NPC_Del
17056010	7	25593351	25750994	*LOC646588*	1.05	5.62E-08	0.00015	NPC_Dup vs NPC_Del
17058645	7	74689057	74728849	*LOC101926943*	0.68	6.46E-08	0.00015	NPC_Dup vs NPC_Del
17058582	7	73736094	73738867	*ABHD11*	0.72	6.13E-08	0.00015	NPC_Dup vs NPC_Del
17046935	7	73433779	73436120	*FZD9*	0.82	6.16E-08	0.00015	NPC_Dup vs NPC_Del
17058557	7	73699205	73719687	*STX1A*	0.59	6.91E-08	0.00015	NPC_Dup vs NPC_Del
17047176	7	74657667	74760696	*GTF2I*	1.16	1.22E-07	0.00025	NPC_Dup vs NPC_Del
17073006	8	136489448	136529876	*RP11-431D12.1*	0.73	1.92E-07	0.00037	NPC_Dup vs NPC_Del
16969766	4	110365733	110565285	*ENPEP*	0.98	2.12E-07	0.00039	NPC_Dup vs NPC_Del
16938468	3	27802762	27891301	*LOC105377007*	0.98	4.50E-07	0.00079	NPC_Dup vs NPC_Del
17097211	9	111549722	111599855	*PTGR1*	-1.29	6.26E-07	0.00105	NPC_Dup vs NPC_Del
16996017	5	45254950	45696151	*HCN1*	1.29	7.85E-07	0.00126	NPC_Dup vs NPC_Del
17058427	7	73302517	73308867	*NSUN5*	0.89	2.42E-06	0.00372	NPC_Dup vs NPC_Del
17092712	9	19108375	19149290	*PLIN2*	-0.95	3.84E-06	0.00568	NPC_Dup vs NPC_Del
17047112	7	74289473	74405943	*CLIP2*	0.63	4.87E-06	0.00692	NPC_Dup vs NPC_Del
16860322	19	23914876	23945159	*ZNF726*	1.01	5.59E-06	0.00765	NPC_Dup vs NPC_Del
17096335	9	97237764	97238678	*ANKRD18CP*	-0.94	6.41E-06	0.00845	NPC_Dup vs NPC_Del
17058617	7	74231499	74254444	*RFC2*	1.01	1.12E-05	0.01430	NPC_Dup vs NPC_Del
17072725	8	128008898	128008956	*MIR1206*	1.85	1.29E-05	0.01583	NPC_Dup vs NPC_Del
16703616	10	29409402	29487745	*SVIL-AS1*	0.77	1.86E-05	0.02212	NPC_Dup vs NPC_Del
16851230	18	14179097	14227050	*ANKRD20A5P*	1.50	2.69E-05	0.03064	NPC_Dup vs NPC_Del
16995873	5	43012698	43067786	*LOC648987*	-0.63	2.82E-05	0.03064	NPC_Dup vs NPC_Del
16999379	5	126541841	126595418	*ALDH7A1*	-0.41	2.79E-05	0.03064	NPC_Dup vs NPC_Del
16950346	3	6631689	6805449	*GRM7-AS3*	-0.46	3.19E-05	0.03363	NPC_Dup vs NPC_Del
17012045	6	117301455	117301758	*RN7SKP51*	-0.54	3.36E-05	0.03445	NPC_Dup vs NPC_Del
16702747	10	16436965	16513737	*PTER*	-0.67	3.74E-05	0.03732	NPC_Dup vs NPC_Del
16854803	18	39678520	39678622	*RNU6-1242P*	-1.37	4.00E-05	0.03885	NPC_Dup vs NPC_Del
17045938	7	48171458	48647496	*ABCA13*	0.64	4.35E-05	0.04117	NPC_Dup vs NPC_Del
17114177	X	131702650	131830643	*FIRRE*	1.36	4.79E-05	0.04424	NPC_Dup vs NPC_Del
16854801	18	39622123	39622175	*MIR924*	-1.59	5.16E-05	0.04645	NPC_Dup vs NPC_Del
16787103	14	82642539	82706825	*LOC101928599*	0.89	5.53E-05	0.04861	NPC_Dup vs NPC_Del
17081546	8	138130023	138496822	*FAM135B*	-1.67	3.18E-10	1.18E-05	Neu_Del vs Neu_Ctrl
16886656	2	154698299	154858351	*KCNJ3*	-1.17	7.20E-09	0.00013	Neu_Del vs Neu_Ctrl
16773596	13	29024611	29505947	*MTUS2*	-1.34	7.82E-07	0.00963	Neu_Del vs Neu_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	-0.92	3.19E-06	0.02942	Neu_Del vs Neu_Ctrl
17106242	X	111511662	111522399	*LINC00890*	1.81	5.43E-06	0.03339	Neu_Del vs Neu_Ctrl
17047138	7	74453790	74602605	*GTF2IRD1*	-0.79	4.82E-06	0.03339	Neu_Del vs Neu_Ctrl
16821092	16	76277278	76559238	*CNTNAP4*	-1.18	8.79E-06	0.04638	Neu_Del vs Neu_Ctrl
17043622	7	8433955	8752963	*NXPH1*	-2.02	8.92E-13	3.30E-08	Neu_Dup vs Neu_Ctrl
16886656	2	154698299	154858351	*KCNJ3*	-1.62	2.13E-12	3.93E-08	Neu_Dup vs Neu_Ctrl
16928510	22	26169474	26383597	*SEZ6L*	-1.89	9.29E-10	1.14E-05	Neu_Dup vs Neu_Ctrl
16705295	10	66926006	67101109	*LRRTM3*	-2.35	2.48E-09	2.29E-05	Neu_Dup vs Neu_Ctrl

(Continued)

AffyID	Chr	Start	End	Symbol	logFC	q-value	q-value	Category
16703382	10	26216307	26304562	*GAD2*	-2.52	6.77E-09	5.00E-05	Neu_Dup vs Neu_Ctrl
16858519	19	11767000	11784078	*ZNF441*	-2.89	3.27E-08	0.00020	Neu_Dup vs Neu_Ctrl
16705305	10	67052609	67055028	*RP11-344A5.1*	-1.03	1.26E-07	0.00066	Neu_Dup vs Neu_Ctrl
16773596	13	29024611	29505947	*MTUS2*	-1.50	1.96E-07	0.00086	Neu_Dup vs Neu_Ctrl
17081546	8	138130023	138496822	*FAM135B*	-1.34	2.10E-07	0.00086	Neu_Dup vs Neu_Ctrl
16981119	4	163524298	164384050	*MARCH1*	-1.10	3.22E-07	0.00119	Neu_Dup vs Neu_Ctrl
16887492	2	170813213	170861151	*GAD1*	-1.81	4.43E-07	0.00136	Neu_Dup vs Neu_Ctrl
16925809	21	40010999	40847139	*DSCAM*	-1.53	4.08E-07	0.00136	Neu_Dup vs Neu_Ctrl
16660436	1	21509365	21578412	*ALPL*	2.13	1.96E-06	0.00557	Neu_Dup vs Neu_Ctrl
16778790	13	45779228	45851736	*SIAH3*	-1.39	5.18E-06	0.01366	Neu_Dup vs Neu_Ctrl
16857248	19	4769105	4772556	*MIR7-3HG*	-1.57	5.61E-06	0.01380	Neu_Dup vs Neu_Ctrl
17052973	7	146116361	148420998	*CNTNAP2*	-1.60	7.62E-06	0.01481	Neu_Dup vs Neu_Ctrl
16913648	20	38724462	38729372	*SLC32A1*	-1.17	7.45E-06	0.01481	Neu_Dup vs Neu_Ctrl
16832234	17	20868497	20998028	*CCDC144NL-AS1*	3.60	7.06E-06	0.01481	Neu_Dup vs Neu_Ctrl
16782710	14	24171853	24180257	*REC8*	-1.94	8.21E-06	0.01515	Neu_Dup vs Neu_Ctrl
16890457	2	214411065	214578959	*VWC2L*	-1.47	8.85E-06	0.01556	Neu_Dup vs Neu_Ctrl
16823799	16	9753404	10182754	*GRIN2A*	-1.49	1.22E-05	0.02051	Neu_Dup vs Neu_Ctrl
17087054	9	94558714	94568030	*PCAT7*	2.52	1.45E-05	0.02322	Neu_Dup vs Neu_Ctrl
16821092	16	76277278	76559238	*CNTNAP4*	-1.19	1.62E-05	0.02460	Neu_Dup vs Neu_Ctrl
16769511	12	105102472	105107179	*LOC414300*	-1.07	1.67E-05	0.02460	Neu_Dup vs Neu_Ctrl
16966497	4	46993723	47426445	*GABRB1*	-1.13	1.91E-05	0.02612	Neu_Dup vs Neu_Ctrl
17114177	X	131702650	131830643	*FIRRE*	2.67	1.87E-05	0.02612	Neu_Dup vs Neu_Ctrl
16666616	1	84078062	84238498	*PRKACB*	-1.51	2.09E-05	0.02756	Neu_Dup vs Neu_Ctrl
16798938	15	32641669	32697098	*SCG5*	-1.47	2.35E-05	0.02901	Neu_Dup vs Neu_Ctrl
16666965	1	89524836	89597864	*LRRC8B*	-1.00	2.36E-05	0.02901	Neu_Dup vs Neu_Ctrl
17053004	7	147671699	147673144	*LOC105375556*	-1.33	2.51E-05	0.02986	Neu_Dup vs Neu_Ctrl
16679033	1	236518000	236552981	*LGALS8*	-1.11	2.83E-05	0.03271	Neu_Dup vs Neu_Ctrl
16890466	2	214510196	214536890	*VWC2L-IT1*	-1.08	3.24E-05	0.03575	Neu_Dup vs Neu_Ctrl
17098866	9	128392618	128392714	*MIR219A2*	2.08	3.29E-05	0.03575	Neu_Dup vs Neu_Ctrl
16934171	22	32149006	32159322	*C22orf42*	-0.81	3.42E-05	0.03607	Neu_Dup vs Neu_Ctrl
16661155	1	26410778	26429728	*LIN28A*	1.88	4.14E-05	0.04247	Neu_Dup vs Neu_Ctrl
16970047	4	118033618	118258648	*NDST3*	-1.12	4.41E-05	0.04404	Neu_Dup vs Neu_Ctrl
17058460	7	73440398	73522285	*BAZ1B*	1.41	2.08E-12	7.68E-08	Neu_Dup vs Neu_Del
17043622	7	8433955	8752963	*NXPH1*	-1.50	3.76E-11	6.94E-07	Neu_Dup vs Neu_Del
16858546	19	11848761	11869491	*ZNF439*	-1.86	3.47E-09	4.08E-05	Neu_Dup vs Neu_Del
17047073	7	74174245	74197101	*EIF4H*	1.33	4.42E-09	4.08E-05	Neu_Dup vs Neu_Del
16959507	3	135261028	135439888	*LOC105374122*	1.50	1.16E-08	8.56E-05	Neu_Dup vs Neu_Del
16858519	19	11767000	11784078	*ZNF441*	-2.45	2.60E-08	0.00016	Neu_Dup vs Neu_Del
17058484	7	73536353	73558002	*BCL7B*	1.11	4.63E-08	0.00024	Neu_Dup vs Neu_Del
17047138	7	74453790	74602605	*GTF2IRD1*	0.79	2.93E-07	0.00135	Neu_Dup vs Neu_Del
17047086	7	74191198	74191294	*MIR590*	2.75	4.38E-07	0.00180	Neu_Dup vs Neu_Del
17046948	7	73683025	73705163	*WBSCR22*	0.56	1.04E-06	0.00383	Neu_Dup vs Neu_Del
16884008	2	107986523	108013994	*SLC5A7*	-1.63	1.52E-06	0.00511	Neu_Dup vs Neu_Del
16681278	1	7843083	7853512	*UTS2*	-1.45	2.55E-06	0.00784	Neu_Dup vs Neu_Del
16928510	22	26169474	26383597	*SEZ6L*	-1.11	2.76E-06	0.00785	Neu_Dup vs Neu_Del
17058499	7	73568944	73578791	*TBL2*	1.11	3.89E-06	0.01027	Neu_Dup vs Neu_Del
16970047	4	118033618	118258648	*NDST3*	-1.08	4.87E-06	0.01198	Neu_Dup vs Neu_Del
16981184	4	166733384	167234796	*SPOCK3*	-0.95	6.62E-06	0.01464	Neu_Dup vs Neu_Del
17068712	8	46831050	46831156	*RNU6-656P*	0.89	6.74E-06	0.01464	Neu_Dup vs Neu_Del
17106242	X	111511662	111522399	*LINC00890*	-1.53	1.07E-05	0.02192	Neu_Dup vs Neu_Del
17106972	X	131058242	131089883	*ARHGAP36*	-1.33	1.24E-05	0.02294	Neu_Dup vs Neu_Del
16886961	2	161424332	161985282	*SLC4A10*	-0.92	1.23E-05	0.02294	Neu_Dup vs Neu_Del
16704822	10	49610301	49612720	*SLC18A3*	-1.41	2.33E-05	0.04105	Neu_Dup vs Neu_Del
16762053	12	21264600	21419594	*SLCO1A2*	1.70	2.56E-05	0.04114	Neu_Dup vs Neu_Del

CHAPTER 3

Transcriptomic analysis unveils consequences of complex rearrangements associated with autism spectrum disorder

Mar Costa-Roger[1], Marcos López-Sánchez[1], Carlos Arenas-Ruiz[1], Teresa Vendrell[2], Patricia Muñoz-Cabello[1,3], Blanca Gener[4], Elisabeth Gabau-Vila[5], Nuria Capdevila[5], Luis A. Pérez-Jurado[1,3], Ivon Cuscó[1,2]

In preparation

ABSTRACT

Introduction. Autism Spectrum Disorder (ASD) is a group of neurodevelopmental disorders characterized by social-communication difficulties, restricted interests and repetitive behaviours. ASD has a complex multifactorial etiology with high genetic involvement. Little is known about complex regions such as segmental duplications, which can predispose to inversions. The aims of this project are to study the association of two common inversions (inv17q21.31/inv8p23.1) with ASD and to elucidate the transcriptional consequences of these rearrangements.

Methods. We performed a transmission disequilibrium test (TDT) using SNP array data from 3,770 ASD trios and 2,217 control trios (n=23,005 individuals) downloaded from public repositories in order to investigate the association between the candidate inversions and ASD. In addition, we carried out an RNAseq analysis of peripheral blood in 13 idiopathic ASD patients and 11 controls grouped by inversion genotypes: non-inverted; inv8p23.1; inv17q21.31. We used HTSeq/DESeq2 to discover differentially expressed genes (DEGs) with q-value<0.05.

Results. We found a significant over-transmission of the inverted alleles of inv17q21.31 (OR=1.12, p=0.0047) and inv8p23.1 (OR=1.12, p=0.0006) in ASD patients from European ancestry. Moreover, our data shows that inversions affect the expression of local genes as well as genome-wide and a great proportion of DEGs correspond to lncRNAs. Also, we identified DEGs associated with ASD and other neuronal diseases involved in immune-inflammatory and actin cytoskeleton pathways in ASD inverted individuals.

Conclusions: TDT suggests that the inverted alleles at the common inversions 17q21.31 and 8p23.1 may act as susceptibility factors for ASD in Europeans. Our transcriptomic results show that inversions have a direct effect on gene expression patterns that could be explained by the disruption of TADs and NHCCs. The fact that we have identified several lncRNAs indicates that gene regulation could have an important role in ASD, particularly implicating *KANSL1-AS1* and *LINC00938*. Moreover, we note that ASD carriers of the susceptibility alleles (inv17q21.31/inv8p23.1), have differential expression of neuronal function-related genes that could be new ASD candidate genes (*ATP13A1*, *NR4A3* and *SLC12A6*). We also highlight the implication of a deregulation of immune-inflammatory pathways and actin cytoskeletal regulation in ASD pathogenesis.

Keywords: inversions; ASD; 8p23.1; 17q21.31; TADs; immune system.

[1]Genetics Unit, Universitat Pompeu Fabra, Hospital del Mar Research Institute (IMIM), Parc de Salut Mar and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Universitat Pompeu Fabra, Departament de Ciències Experimentals i de la Salut, Barcelona, Spain

[2]Genetics Service, Parc de Salut Mar, Barcelona, Spain

[3]Department of Clinical and Molecular Genetics and Medicine Genetics Group, VHIR, University Hospital Vall d'Hebron, Barcelona, Spain

[4]Osakidetza Basque Health Service, Cruces University Hospital. Department of Genetics, Biocruces Bizkaia Health Research Institute, Bizkaia, Spain

[5]Service of Clinical Genetics and Genetics Laboratory of Paediatric Unit, Parc Taulí University Hospital, Institut d'Investigació I Innovació Parc Taulí I3PT, Universitat Autònoma de Barcelona, Sabadell, Spain

INTRODUCTION

Autism Spectrum Disorder (ASD) is a heterogeneous neurodevelopmental disorder characterized by impairments in social communication and interaction as well as restricted and repetitive behaviours or interests (1). It occurs with a population prevalence of ~1.5% in developed countries with a male-to-female ratio of 3-4:1, which poses an important health problem worldwide (2,3). Twin studies have enabled to assess the proportion of the phenotype variance accounted for genetic factors, concluding that ASD is highly heritable. Recent studies estimated ASD heritability to be around 83-85% (4,5). Although it is known that ASD has a complex multifactorial etiology, this data demonstrates that genetic factors play a crucial role.

Technological advances have allowed an enormous progress in regard with the genetic architecture of ASD highlighting a large number of genes involved in this pathology (6–8). However, only in ~25-35% of patients diagnosed with ASD, the genetic etiology is determined. The most common causes identified are monogenic syndromes associated with ASD, copy number variants and single nucleotide variants. This is why ASD is considered genetically complex existing highly penetrant monogenic cases, polygenic models and multifactorial models involving an environment-genome interaction (reviewed by 9).

Single copy regions and coding regions have been exhaustively studied in ASD patients, but little is known about complex regions and non-coding regions. Segmental duplications (SDs) are part of these unexplored regions, which may contain part of the lost heritability. SDs, also known as low-copy repeats (LCRs), are segments of DNA between 10-300 Kb in length with high homology (>95%) and found in multiple genomic locations. The human SD content is estimated to be around ~5.5% (10,11). Due to their similarity, they predispose to genomic instability mediated by non-allelic homologous recombination (NAHR) events that lead to recurrent rearrangements including duplications, deletions, inversions, marker chromosomes or translocations (12,13).

Inversions are known to contribute to chromosomal evolution and are considered a source of genetic variability. Although inversions are usually considered neutral variants, inversion breakpoints can cause gene disruption, alter gene expression due to separation of regulatory elements and predispose to genetic copy number imbalances (14). These rearrangements have been linked to different diseases, including multifactorial pathologies such as autism and obesity (15–17). The candidate inversions, inv17q21.31 and inv8p23.1, have been selected because both are well-studied and validated ancestral inversions, which allow its genotyping with *scoreInvHap* using SNP (Single nucleotide Polymorphism) data. This tool contains a set of experimentally validated reference haplotype-genotypes from the inversions of interest and classifies individuals according to their SNP similarities (18). In fact, transcriptome-wide significant associations have been reported in both regions in an ASD study integrating expression data with GWAS (19).

The inversion 17q21.31 is a ~900 Kb inversion with two highly divergent haplotype groups (H1 – non-inverted / H2 – inverted) and mostly present in European population with a frequency of ~20% (20,21). This region contains different genes including *MAPT* (Microtubule-associated protein tau) and *CRHR1* (Corticotrophin releasing hormone receptor 1). De Jong *et al.* described altered gene expression including overexpression of *LRRC37A* and *CRHR1* as well as decreased expression of *LRRC37A4, PLEKHM1* and *MAPT* in H2 in several tissues (22). The H1 haplotype is associated to an increased *MAPT* expression, which produces an overproduction of hyperphosphorylated protein Tau that aggregates in neurons. This mechanism has been linked to neurodegenerative disorders including progressive supranuclear palsy, corticobasal degeneration, Parkinson's disease and Alzheimer's disease (23–27). In addition, the H2 haplotype is related to a recurrent microdeletion predisposing to the 17q21.31 microdeletion syndrome, which is characterized by developmental delay and learning disabilities (28).

The 8p23.1 inversion is the largest inversion found in the human genome including almost 4.7 Mb with estimated frequencies depending on the

population; 59-80% in Africans, 13-33% in Asians and 29-50% in Europeans (20,29–31). Genes encompassing this inversion, such as *PPP1R3B* (Protein phosphatase 1 regulatory subunit 3B), have been related to lipid metabolism (32,33) and an association was found between the 8p23 inverted allele (I) and low BMI in children of European ancestry (34). Also, variants in *FAM167A/BLK* locus, harboured in the inversion region, have been linked to risk to systemic lupus erythematosus (SLE) and other autoimmune diseases (35–38). Recent studies detected an association effect between the non-inverted status of inv8p23.1 and susceptibility to SLE (OR=1.18, $p=8.18 \times 10^{-7}$) (39,40).

In the present work, we investigated the association of inv17q21.31 and inv8p23.1 with ASD through TDT (transmission disequilibrium test). We have also performed transcriptomic studies in order to elucidate the gene expression pattern of these inversions, and to understand which mechanisms and pathways could be implicated in the ASD phenotype.

METHODS

Datasets for TDT

The association of inv17q21.31 and inv8p23.1 was studied in six datasets, three ASD datasets (AGP, SSC and UMSGARD) and three control datasets (MCTFR, IMAGE and SHARP). We were granted permission from dbGaP (Database of Genotypes and Phenotypes, http://www.ncbi.nlm.nih.gov/gap) and the Simons Foundation Autism Research Initiative (SFARI) to download genotypes and clinical information.

Autism Genome Project (AGP) (dbGaP phs000267.v4.p2) includes genotype data of a total of 2,562 parent-offspring trios performed with Illumina Human 1M-duo Infinium BeadChip SNP array. Simons Simplex Collection (SSC, http://sfari.org/resources/simons-simplex-collection) consists of 2,605 families genotyped with Illumina arrays with more than 1 million SNPs including 1Mv1, 1Mv3 or Omni2.5. Each family was composed of trios or quartets with a non-affected sibling available in 2,433 of the cases. Probands from AGP and SSC

were diagnosed using ADI-R and ADOS instruments. University of Miami Study on Genetics of Autism and Related Disorders (UMSGARD) (dbGaP phs000436.v1.p1) comprises genotype data of 422 unrelated parent-child trios. Probands included in the study have a clinical diagnosis of ASD and are from the University of Miami study on autism disorders.

Regarding controls, we used three independent trio datasets. Minnesota Center for Twin and Family Research (MCTFR) Genome-Wide Association Study of Behavioural Disinhibition (dbGaP phs000620.v1.p1) includes genotype data of 1,478 families that formed part of the MCTFR longitudinal study. Individuals were genotyped with the Illumina platform Human660W-Quad v1.0. International Multi-Center ADHD (Attention Deficit Hyperactivity Disorder) Genetics Project (IMAGE) (dbGaP phs000016.vs.p2) consists of 958 parent-child trios genotyped with Perlen-600K array. Probands were children aged 6-17 years with IQ above 70, any subtype of ADHD, single-gene disorders or neurological disease and not meeting criteria for ASD. SNP Health Asthma Resource Project (SHARP) (dbGaP phs000166.v2.p1) comprises genotype data of a total of 435 parent-offspring trios obtained with Affymetrix 6.0 platform.

European ancestry assessment

Although TDT is not sensitive to population stratification, heterogeneity in ancestry could dilute the signal of a geographically restricted segregation. Therefore, TDT was performed only in European samples. In order to determine European ancestry of samples, we used the software peddy (41). We selected those individuals identified as European by peddy.

SNP imputation and inversion genotyping

SNP microarray data were imputed using the imputeInversion pipeline prior to inversion calling as performed in González JR, et al. (2020) (17). This pipeline was designed to impute only those SNPs inside the inversion region or closer than 500 kb to inversion breakpoints. imputeInversion uses shapeITv2.r904 to phase, Minimac3 to impute, and 1000 Genomes as reference haplotypes. Inversion genotyping was performed with the package *scoreInvHap* (18), which

uses SNP data to compute similarity scores between the genotypes of an individual and the reference genotypes for the 3 inversion status.

Genetic association analyses

For TDT, we analysed only complete trios (parents and probands and/or sibling if available). After performing imputation and European ancestry assessment, we obtained a total of 3,770 ASD trios and 2,217 control trios (n=23,005 individuals). TDT is a McNemar's test based on binomial distribution, which we carried out using PLINK (42), an open-source whole genome association analysis toolset. We performed TDT in all datasets for all parental transmissions, separately analysing paternal and maternal transmissions. We considered loci to be over-transmitted when reached a particular significance threshold ($p<0.05$), and parental over-transmission occurred when the significance threshold was only derived from maternal or paternal over-transmission. A standard Chi-squared (χ^2) test was used to evaluate the transmission ratio, and the goodness-of-fit χ^2 test to determine if paternal and maternal transmission ratio deviated from the expected ratio.

Selection of subjects for the transcriptomic study

In our laboratory, we have gathered a total of 112 ASD patients (91 males, 21 females) referred from several hospitals of the Spanish Health System. All individuals had a diagnosis of ASD based on clinical and/or psychological evaluations, and underwent a detailed clinical examination as well as family history review to search for syndromic forms of autism that were discarded. All cases are idiopathic with negative karyotype, Fragile X test, molecular karyotype and exome. Individuals were homogeneous regarding age (<25 years old). Written informed consent from parents or legal caregivers was obtained for all cases and the study was approved by the Clinical Research Ethics Committee (CEIC) of the Parc de Salut MAR.

In order to study the transcriptomic consequences of inv17q21.31 and inv8p23.1, we selected unrelated patients according to their inversions genotypes (13 males). As controls we included 11 adult individuals of Spanish

origin (4 males, 7 females). These individuals were classified according to their inversion genotype in three groups: (1) 5 ASD patients and 3 controls non-inverted for both regions (H1/H1 inv17q21.31 and NI/NI inv8p23.1); (2) 4 ASD subjects and 4 controls inverted for 17q21.31 (H2/H2 inv17q21.31 and NI/I 8p23.1); (3) 4 ASD patients and 4 controls inverted for 8p23.1 (H1/H1 inv17q21.31 and I/I inv8p23.1). See table from Figure 1 and Supplementary Table 1.

RNA extraction, quality and quantification

From fresh whole blood samples, peripheral mononuclear cells (PBMCs) were isolated by ficoll density gradient centrifugation method (Lymphoprep™ STEMCELL Technologies). Total RNA was extracted using the RNeasy Mini kit (Qiagen) following manufacturer's instructions or Trizol (Life Technologies) following standard protocol. After RNA extraction with Trizol, samples were purified using the Qiagen column in order to improve the quality. Quantification and quality of RNA samples was assessed using a NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer. RNA integrity number (RIN) values ranged from 6.5 to 10.

Transcriptome sequencing

The RNAseq study was conducted on a HiSeq 4000 (Illumina) and paired-end sequences were acquired at a read length of 75 bp with 25,368,875 mean read pairs per sample. Quality control analysis of raw data was done using FastQC software and sequence reads were aligned to the Genome Reference Consortium Human genome build 37 (GRCh37) using STAR aligner (43). Gene expression levels were quantified by htseq-count from the HTSeq python package (44) and 'DESeq2' R package (v3.10) (45) was used to assess differentially expressed genes (DEGs) between different sample groups. In the design formula of DESeq2, we considered corrections by sex and RNA extraction procedure in order to minimise possible biases (Supplementary Table 1). The non-inverted state was considered the reference state in both inversions. Wald-log test was applied and p-values were adjusted for False Discovery Rate (FDR) according to Benjamini-Hochberg correction procedure. Differential expression

quantification was expressed in Log2 fold change (log_2FC) and DEGs with q-value<0.05 were considered statistically significant. We used two different approaches for each inversion. In the first one, only the inversion status was considered, regardless of the subjects' phenotype. In the second approach, inverted ASD individuals were compared against all other groups in order to elucidate unique gene signatures from inverted ASD patients. A schematic representation of the strategy followed is depicted in figure 1.

Pathway enrichment analysis

To elucidate deregulated mechanisms affected by DEGs, pathway enrichment analysis was performed using the publicly available Consensus Path database (CPDB) (46), which integrates interaction networks from 32 public resources. Statistical analyses were done with the CPDB overrepresentation analysis option including pathway-based and Gene Ontology (GO)-based tests. The q-value threshold of 0.05 was used.

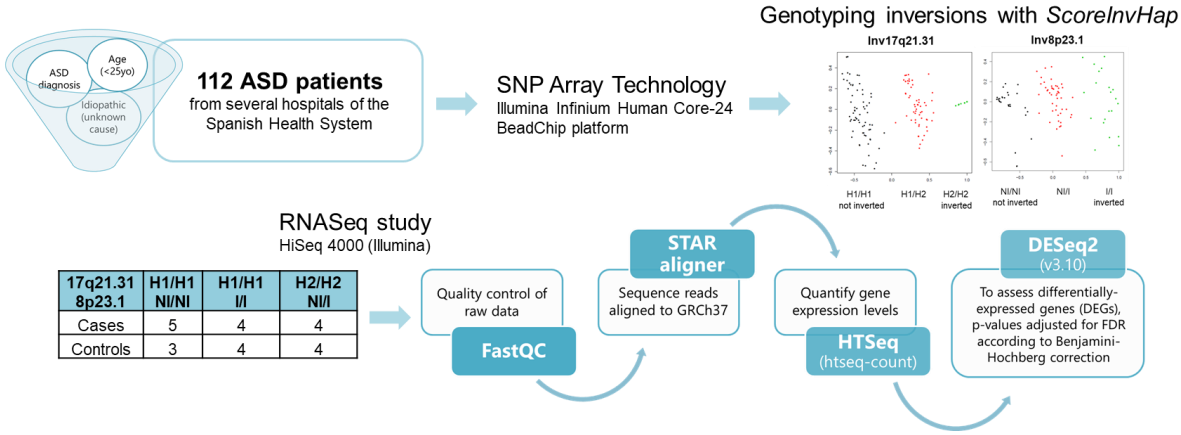


Figure 1. Strategy followed for the transcriptomic study. From the 112 ASD individuals, we selected 13 patients to perform a RNAseq study depending on their inversion status genotyped with *scoreInvHap*.

RESULTS

TDT shows over-transmission of inverted alleles for inv17q21.31 and for inv8p23.1 in ASD probands

The inverted allele H2 at inv17q21.31 was found significantly over-transmitted in ASD probands of the AGP trio dataset (OT=9.04, OR=1.20, p=0.00094) (Table 1 and Supplementary Table 2). We analysed two additional datasets with ASD trios (SSC and UMSGARD). The association of ASD with over-transmission of the H2 haplotype was not replicated in SSC (OT=1.78, OR=1.04, p=0.5508). In UMSGARD, the over-transmission of the inverted allele was also noticed, although it did not reach significance likely due to the small sample size (OT=2.86, OR=1.06, p=0.6788). A meta-analysis of the three datasets provided significant evidence of an over-transmission of the H2 allele in ASD probands (OR=1.12, p=0.0047), considering only families of European ancestry (3,770 trios) (Table 1). Given the known phenotypic heterogeneity of ASD, we also performed a stratified analysis by the different clinical variables or subphenotypes. The most significant over-transmission occurred in multiplex families as well as in ASD probands who did not fulfill criteria for strict autism and with intellectual quotient (IQ) above 80 (Supplementary Table 3). Maternal and paternal-specific over-transmission did not show significant results. However, meta-analysis of non-ASD trios (MCTFR, IMAGE and SHARP; 2,217 trios) revealed a non-significant tendency for H2 allele of inv17q21.31 (OR=1.07, p=0.2233), showing some transmission distortion to a less extent in the general population of the same allele associated with ASD risk.

Over-transmission of the inverted I-allele at inv8p23.1 was observed consistently in the three independent ASD datasets, significant in AGP ((OT=6.88, OR=1.15, p=0.0036) and with the same tendency in SSC (OT=4.36, OR=1.09, p=0.0829) and UMSGARD (OT=5.23, OR=1.11, p=0.3604). A meta-analysis of the three studies provided significant evidence of an over-transmission of the inverted allele at inv8p23.1 (OR=1.12, p=0.0006) in European samples (Table 1). Stratification analysis using subphenotypes only yielded significant increment of the association for verbal ASD and non-strict

autism (Supplementary Table 3). In this case, the maternal and paternal transmission was also similar with no gender bias. In non-ASD trios (MCTFR, IMAGE and SHARP), the meta-analysis did not show over-transmission of any allele at inv8p23.1 (OR=1.00, p=0.9496).

		inv17q21.31				inv8p23.1			
	Trios	**ME**	**%OT**	**OR**	**p-value**	**ME**	**%OT**	**OR**	**p-value**
AGP	1868	0	9.04	1.20	0.0009	15	6.88	1.15	0.0036
SSC	1613	0	1.78	1.04	0.5508	18	4.36	1.09	0.0829
UMSGARD	289	0	2.86	1.06	0.6788	1	5.23	1.11	0.3604
ASD Eur (all)	3770	**0**		**1.12**	**0.0047**	**34**		**1.12**	**0.0006**
Controls Eur	2217	0		1.07	0.2233	21		1.00	0.9496

Table 1. TDT meta-analysis of ASD and control trios from multiple datasets for inv17q21.31 and inv8p23.1. Trios: number of complete trios with genotype data used, ME: Mendelian errors, OT: over-transmission (%), OR: odds ratio.

Gene expression pattern in inv17q21.31 and inv8p23.1

In order to characterize the functional consequences of the inversion polymorphisms, we performed a first approach where we compared mRNA levels of non-inverted (reference level) with inverted subjects.

For inv17q21.31, our results revealed a total of 38 differentially expressed genes (DEGs). We did not detect expression of *CRHR1* gene, as it is not expressed in blood, whereas *MAPT* showed a downregulated tendency in H2 without reaching statistical significance as it was expected (22). A considerable proportion of DEGs was located in chromosome 17 (12/38, 31.5%) and concretely, within the 17q21.31 inverted loci, we found 8 DEGs, including single-copy region and multiple-copy genes from SDs (Figure 2A). The expression pattern was in agreement with the expression of tagSNPs (rs241027) from GTEx (Genotype-Tissue Expression) in whole blood and lymphocytes (47). This region flanked by SDs contains topologically associated domains (TADs), which are clusters of genomic interactions of 100 kb to 1 Mb in size found in mammalian genomes (figure 2B) (48). The fact that the inversion breakpoints could modify the 3D structure of the TADs could be a possible explanation for affecting the gene expression of those genes located in the single-copy region.

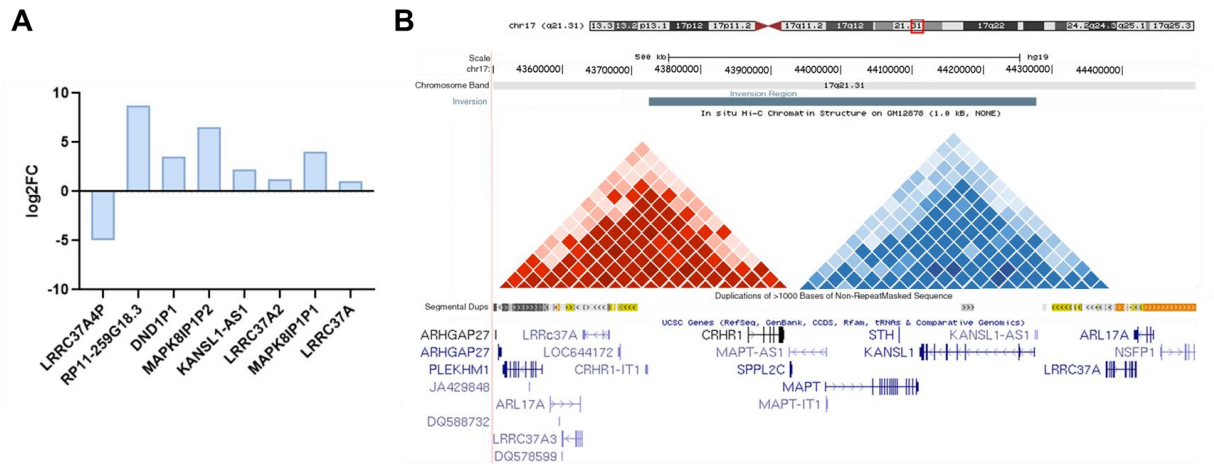


Figure 2. Differentially expressed genes in the 17q21.31 inverted region. (a) Gene expression patterns of DEGs located in the region. (b) Schematic representation of the 17q21.31 region with the TADs present at this loci.

For inv8p23.1, we only detected *FAM167A* to be downregulated ($log_2FC=-1.72$; q-value=0.0133) in individuals carrying the I-alleles compared with non-inverted individuals. As in the 17q21.31 loci, the 8p23.1 region is surrounded by TADs (Supplementary Figure 1), which could be disrupted by inversion breakpoints possibly accounting for the change in expression of single-copy genes detected.

Specific ASD transcriptional consequences in candidate inversions

Once proven that inversions directly affect gene expression, differential gene expression was analysed taking into account phenotype subgroups in order to elucidate unique ASD transcriptional consequences in both candidate inversions.

Regarding inv17q21.31, from the 38 DEGs previously identified, we found 4 that were significantly different in ASD patients carrying the 17q21.31 inverted alleles, including 2 from the 17q21.31 inverted interval (*KANSL1-AS1* and *LRRC37A4P*) and 2 from elsewhere in the genome (*IL1B* and *CCL3*). In particular, *KANSL1-AS1* was upregulated in H2-I allele carriers ASD patients compared to inverted controls (q-value=0.0029) and to the non-inverted individuals (q-value<0.0001). Controls carrying the inversion showed an intermediate

expression, while H1 group had the lowest one (q-value=0.0511) (Figure 3A). For *LRCC37A4P*, we observed the opposite situation where ASD inverted patients showed the gene downregulated compared to non-inverted individuals (q-value=0.0002) and to inverted controls, although it was not significant (q-value=0.995) (Figure 3B).

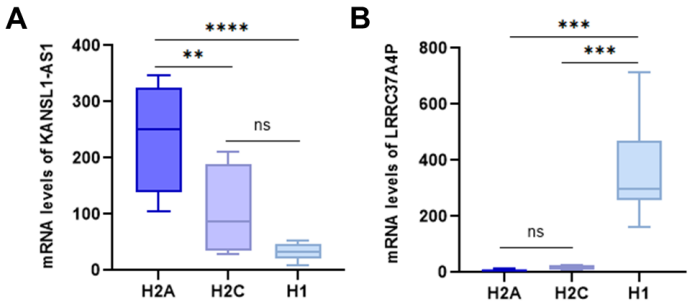


Figure 3. Differentially expressed genes between inverted individuals (H2/H2) and non-inverted (H1/H1) for 17q21.31 region. (a) Differentially expressed genes in the 17q21.31 region when comparing H2 and H1 individuals regardless of their phenotype. (b) mRNA levels of KANSL1-AS1. (c) mRNA levels of LRRC37A4P. P values are shown with asterisks indicating values that are significantly different in a one-way ANOVA (ns: not significant, **: p<0.01, ***: p<0.001, ****: p<0.0001).

Apart from the 4 mentioned genes, we also detected 15 DEGs in this second approach obtaining a total of 19 differentially expressed genes, including 12 downregulated and 7 upregulated (Figure 4 and Supplementary Table 4). The relative expression of detected genes was represented in a heatmap (Figure 4A). These genes were located mainly among chromosomes 1, 2, and 17. We found the same number of protein coding genes than lncRNAs (36.84%), four pseudogenes (21.05%) and one snoRNA (5.26%). See figure 4B and 4C. Remarkably, *PTGS2* has been previously associated with ASD according to SFARI database and *RHOB* has been related to synaptic plasticity.

In order to elucidate the deregulated pathways affected by the unique DEGs detected in inverted ASD individuals, we performed an enrichment pathway analysis. We identified an enrichment for pathways (16/29) and GO terms (39/219) related to the immune system and inflammatory processes (Figure 4D). The top significant pathways included Interleukin-10 signaling, cytokine-cytokine receptor interaction and Rheumatoid arthritis. Interestingly, the

majority of DEGs involved in these pathways were upregulated in our analysis, pointing out an upregulation of immune functions among the detected DEGs.

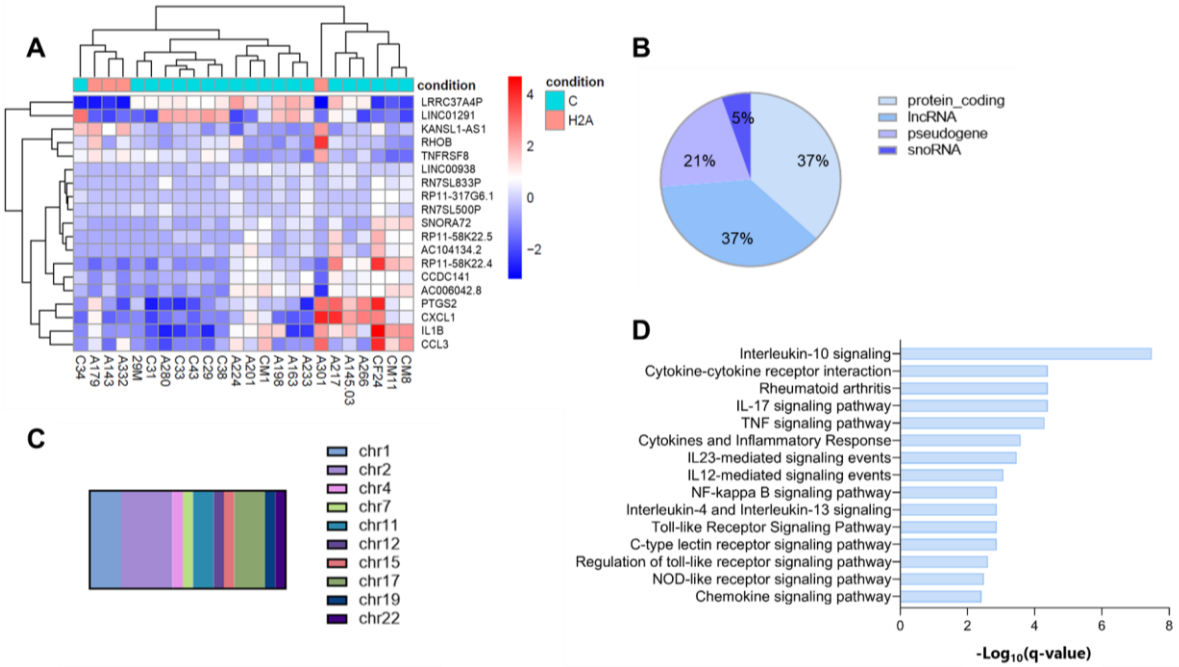


Figure 4. Differentially expressed genes between inverted ASD individuals and all other groups for the 17q21.31 inversion. (a) Heatmap showing the individual expression of each DEG. Blue indicates downregulation, whereas red means upregulation. (b) Biotype of the detected DEGs. (c) Chromosomal distribution of the genes identified. (d) Immune system and inflammatory pathways identified by the enrichment analysis.

For inv8p23.1, when adjusting for phenotype, we detected a total of 36 genome-wide DEGs, including 10 downregulated and 26 upregulated genes (Figure 5 and Supplementary Table 5). Among them, *FAM167A* was found downregulated, but the effect was higher in I/I ASD patients than in inverted controls and non-inverted individuals (Figure 5A).

From the 8p23.1 inverted interval, we also identified decreased expression of *NEIL2* (log_2FC=-0.60; q-value=0.0486) among ASD patients carrying the I-allele. We detected 20 protein coding genes (55.55%), 10 lncRNAs (27.78%), five pseudogenes (13.88%) and one miRNA (2.78%). DEGs were mostly distributed in chromosome 17, but a significant number was located in chromosomes 2, 6, 8, 10, 14 and 19. See figure 5B and 5C. The relative expression of each of the 36

DEGs was represented in a heatmap (Figure 5D). We discovered genes previously reported in ASD literature (*MYH10*), associated with neuronal pathologies (*NR4A3, ATP13A1* and *SLC12A6*) as well as genes with brain-related functions (*WDR47, NSF* and *PDZD8*). In the enrichment analysis, we did not detect any deregulated pathway, but interestingly, we identified GO terms related to actin filament binding and cortical cytoskeleton (figure 5E).

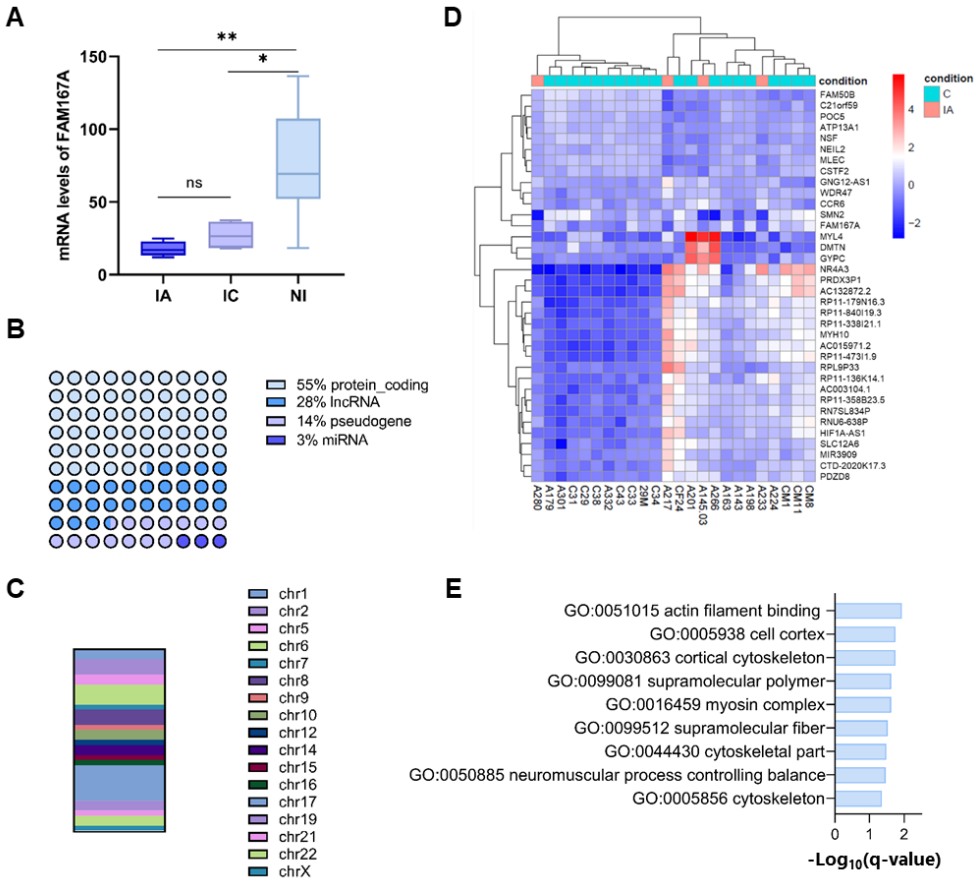


Figure 5. Differentially expressed genes between I/I ASD individuals and all other groups for inv8p23.1. (a) mRNA levels of *FAM167A*. P values are shown with asterisks indicating values that are significantly different in a one-way ANOVA (ns: not significant, *: $p<0.05$; **: $p<0.01$). (b) Biotype of the DEGs and (c) chromosomal distribution. (d) Heatmap showing the downregulated genes indicated with blue colour, and the upregulated in red. (e) GO categories detected in the enrichment analysis.

DISCUSSION

Complex genomic regions containing common inversions could act as susceptibility factors in complex disorders as is the case of the common 16p11.2 inversion which provides a genetic basis for the joint susceptibility to asthma and obesity in European populations (16,17). Our TDT studies showed a significant over-transmission of the inverted H2 allele at inv17q21.31 in ASD patients. However, we observed remarkable differences between studies with completely null results in SSC. A possible explanation for the discrepancy can be related to the composition and phenotype of the samples in the ASD datasets. While AGP and UMSGARD are enriched in multiplex families and include a significant proportion of ASD cases that do not fulfill criteria for strict autism, SSC is restricted to simplex families with the majority of cases with strict autism. Thus, the H2 allele at inv17q21.31 appears to be associated with increased risk for a specific subphenotype of ASD with less strict autism and high IQ, and is more prevalent in families with more than one case of ASD. Interestingly, the risk allele for ASD is also being over-transmitted in the European general population, suggesting that inv17q21.31 is undergoing adaptive selection as previously reported in a study with Icelanders where carrier females had more children and higher genome-wide recombination rates than non-carriers (21). Regarding inv8p23.1, our TDT showed a significant over-transmission of the inverted allele in ASD patients. Therefore, our results revealed that inverted alleles at the common inversions 17q21.31 and 8p23.1 could act as susceptibility risk factors for ASD in Europeans explaining part of the missing heritability in ASD liability.

Focusing on transcriptomic analyses, our aim in the first approach, comparing inverted vs non-inverted individuals, was to further characterize the local gene expression of candidate inversions. We found DEGs within the inverted region both in inv17q21.31 (such as *KANSL1-AS1* and *DND1P1*) and inv8p23.1 (*FAM167A*). Our findings are in line with previously reported associations between H2-I allele carriers of inv17q21.31 and upregulation of *LRRC37A* in brain and decreased levels of *LRRC37A4* in whole blood as well as the gene

expression pattern from GTEx (22,47). Conversely, transcriptomic data from lymphoblastoid cell lines and post-mortem liver samples showed an overexpression of *FAM167A* related to the inverted status of inv8p23.1, which is in the opposite direction compared to our results (29).

The fact that we have identified DEGs within the inverted regions suggests that the expression of these genes is directly affected by the presence of the inversions. TADs are clusters of highly self-interacting regions in the genome where chromosomes form intra-molecularly organized structures allowing regulatory elements to interact with nearby genes (48). Interestingly, there are TADs in both inv17q21.31 and inv8p23.1 boundaries (49). Structural genomic variants, including deletions, duplications, insertions, translocations or inversions, can modify the 3D organization of the genome by disrupting TADs and cause regulatory gain and/or loss of function affecting gene expression (50,51). In some cases, it can be the cause of disease phenotypes as the duplication of a TAD boundary at the *SOX9* locus resulting in Cooks syndrome (52). In fact, around 7% of balanced anomalies cause TAD disruption in developmental disorders (53). Also, rare structural variants disrupting cis-regulatory elements (CRE-SVs) have shown to confer risk for ASD (54). Specifically, inversions are only suspected to influence gene regulation when disrupting a coding gene or TAD boundary (51). In our case, TADs are present in both inversions boundaries and they seem to be disrupted by inversion breakpoints, suggesting that the change in gene expression can be explained by a direct inversion-mediated mechanism. However, chromosome conformation capture techniques, such as Hi-C, should be performed in order to experimentally confirm these results.

However, the majority of significantly affected genes in our study are not located in the inverted regions, but elsewhere in the genome. We speculate disruption of non-homologous chromosomal contacts (NHCCs) as a possible mechanism to explain alteration of these genes. In 1885, the first notions of interchromosomal contacts were described when Carl Rabl proposed that chromosomes occupy defined chromosomal territories in the nucleus in

interphase and that they could interact with nearby chromosomes (55). Non-randomly organized chromosomal territories of interphase chromosomes intermingle to form NHCCs (56). The most well-studied NHCC event is the formation of the nucleolus, where around 300 ribosomal genes located on 5 different chromosomes converge in the same interchromosomal space (57). Therefore, in NHCCs gene regulatory regions from one chromosome can affect genes located in other chromosomes. Despite the lack of information regarding the NHCCs mechanism, it is known that lncRNAs are involved in 3D organization of NHCCs (50). In addition, a study described a high interaction frequency between chromosomes 14-17 and 19-22 (58). These findings lend support to our results, which showed a remarkable number of DEGs located in chromosomes 14 and 17 as well as many deregulated lncRNAs. Moreover, structural and numerical chromosomal rearrangements can cause disruption and reorganization of the network of NHCCs (50). In fact, a recent study analysing chromosomal contacts of CNVs at 16p11.2 locus discovered *cis*- and *trans*-chromatin contacts involving known ASD candidate genes from other chromosomes. It is suggested that disruption of these contacts could influence the observed phenotypes including autism, BMI (body mass index) and head circumference (59).

In the second approach, when adjusting for phenotype subgroups, our results showed that ASD inverted individuals have unique transcriptomic signatures that could be important for ASD phenotype. We have identified a great proportion of deregulated lncRNAs as an ASD specific signature for both candidate inversions representing 37% of DEGs in inv17q21.31 and 28% in inv8p23.1. LncRNAs are transcripts of >200 nucleotides that do not encode for proteins with high expression in the central nervous system. They have key roles in brain development and have been involved in gene expression regulation at different levels including transcriptional, post-transcriptional and epigenetic regulation (60,61). Transcriptomic analysis of post-mortem brain tissues from ASD patients identified 222 differentially expressed lncRNAs (62). Similarly, altered lncRNAs levels have been detected in both ASD brains and peripheral blood (63–65). In addition, a study detected a genome-wide significant ASD-

associated SNP by GWAS, which was located on 5p14.1 and mapped to a lncRNA named *MSNP1AS* (66). Other lncRNAs have also been related to ASD including *HTR5A-AS1* (67), *RPS10P2-AS1* (68), *LINC00693* and *LINC00689* (64), among others. The fact that lncRNAs have been implicated in synaptogenesis, neurogenesis and GABAergic interneuron functions in brain development might explain the functional relation between ASD and lncRNAs, as these functions have also been found deregulated in ASD probands (61). The lncRNAs detected in our analysis did not overlap with lncRNAs previously reported in other studies. We also looked at which tissues were expressed and as lncRNAs can have *cis*-regulatory mechanisms, we identified the nearest neighbouring genes. Therefore, we propose *KANSL1-AS1* and *LINC00938* as new candidate lncRNAs that could have a potential role in ASD, as both have expression in the brain and the nearest neighbouring genes are reported in SFARI (*KANSL1* and *ARID2*). Taken together, our results support an implication of lncRNAs in ASD and in regulation of gene expression.

Regarding protein-coding genes, we found several deregulated genes with neuronal and brain-related functions including *PTGS2, IL1B, CCL3* and *RHOB* for inv17q21.31 and *MYH10, ATP13A1, NR4A3, SLC12A6, PDZD8* and *WDR47* for inv8p23.1. *PTGS2* (Prostaglandin-Endoperoxidase synthase 2) is the key enzyme for prostaglandin biosynthesis with a role in neuronal secretion during neuroinflammation and polymorphisms in this gene have been previously related to ASD (69). Interleukin 1 Beta (*IL1B)* encodes for a proinflammatory cytokine widely expressed in brain that modulates synaptic plasticity and mediates neuroimmune responses to disease and neurodegeneration (70). In our study, *IL1B* was significantly upregulated in ASD subjects carrying the 17q21.31 inverted alleles. In fact, elevated IL-1β serum levels were reported in ASD patients (71) and polymorphisms in this gene have been associated with increased autism susceptibility (71,72). Moreover, several genetic variants in *IL1B* gene have also been linked to schizophrenia and bipolar disorder (73). *CCL3* (C-C motif chemokine ligand 3) and *RHOB* (Ras homology family member B) have been implicated in neuroinflammation processes and in dendritic and spine morphology, respectively (74,75). *MYH10* (Myosin heavy chain 10) is a

strong ASD candidate according to SFARI database as *de novo* rare variants and mosaic variants have been observed in ASD patients (76,77). ATPase 13A1 (*ATP13A1)* expression is increased at the peak of neurogenesis in mice development and recently has been proposed as a novel candidate for ID (78,79). Among other functions, *NR4A3* (Nuclear receptor subfamily 4 group A member 3) is thought to mediate neuronal survival and axon guidance in hippocampal development and has been involved in the nicotine addiction observed in schizophrenia and bipolar disorder patients (80). *SLC12A6* (Solute carrier family 12 member 6) is located in 15q13-14 region, a susceptibility locus for bipolar disorder, schizophrenia, epilepsy, and others neuropsychiatric disorders (81,82). Rare variants in this gene have been associated with bipolar disorder and schizophrenia (83). In addition, recessive mutations in *SLC12A6* are the cause of Andermann syndrome (OMIM #218000), characterized by agenesis of the corpus callosum with severe peripheral neuropathy and psychoses (84). We also highlight *PDZD8* (PDZ domain containing 8) and *WDR47* (WD repeat domain 47) involved in mitochondrial calcium (Ca^{2+}) uptake in neurons and neuronal polarity, respectively (85,86). Therefore, we support the role of *PTGS2, MYH10* and *IL1B* in ASD pathogenesis and we propose the other mentioned genes as new ASD candidate risk genes. As particularly compelling ASD candidates would be those genes linked to other neuropsychiatric disorders such as *ATP13A1, NR4A3* and *SLC12A6* as these disorders and ASD share common genetic background (87,88).

DEGs from 17q21.31 inverted ASD individuals were enriched for pathways and GO terms related to immune system and inflammatory processes, mainly involving upregulated DEGs. Our results support the implication of a deregulation of immune-inflammatory pathways in ASD pathogenesis as it has been previously described. Prior RNAseq analyses from ASD post-mortem brain tissues found an enrichment for immune and inflammatory response in upregulated modules (64,89,90), but findings in blood microarrays are controversial. First studies found upregulation of several pathways linked to immune and inflammatory response in ASD such as IL-2 signaling or leukocyte function as well as upregulation of pro-inflammatory cytokines including *PTGS2*

(91–93). However, other papers reported both upregulated and downregulated genes implicated in immune-inflammatory pathways (94–96). Contrarily, a mega-analysis of ASD blood transcriptomes observed an under-expression of genes with innate and adaptive immunity (97), and a recent RNAseq analysis from discordant sibling pairs was also characterized by decreased peripheral immune-inflammatory functions in ASD (98). Multiple immune cell types have been found deregulated in peripheral blood and brain from ASD patients compared to healthy controls, including altered cytokine and chemokine profiles such as IL-4, IL-10, IL-13 and IL-17, among others (99–101). In fact, the signaling pathways of these concrete interleukins, together with some others, were among the enriched pathways detected in our analysis. Other evidence supporting immune-inflammatory abnormalities in ASD are the higher frequency of infections or allergies displayed by ASD patients, and the fact that maternal infections and maternal immune system vulnerabilities during pregnancy are considered ASD risk factors (99,100). In addition, pro-inflammatory cytokine levels and sustained microglial activation in post-mortem brain tissues indicated a strong neuroinflammation state associated with ASD that can lead to neurodestructive effects (99,102). Immune system is known to play important roles in neurodevelopment and normal brain function such as synaptic plasticity or synaptic pruning, therefore immune abnormalities could alter neuronal synaptic function (103). Although evidence points towards a role of immune dysfunction in the pathogenesis of ASD, more research is needed to elucidate if the immune abnormalities have a causative or a consequential role.

Regarding inv8p23.1, some of the GO terms in ASD inverted individuals were related to actin filament binding and cytoskeleton. Interestingly, regulation of cytoskeletal dynamics is essential for dendritic spine morphology and formation (104) and its alteration has been described in ASD. In particular, several ASD risk genes identified by genetic studies and gene expression studies are involved in cytoskeleton regulation (105). Other evidences include an autism-like mouse model (Shank3-deficient mouse) which experienced a marked loss of cortical actin filaments (106), and a recent study that found altered dynamics of actin

filaments reconstruction in 53% of stem cells derived from ASD patients (107). Our data also supports an impaired function of actin cytoskeleton regulation.

To sum up, TDT suggests that the inverted alleles at inv17q21.31 and inv8p23.1 may act as susceptibility factors for ASD in Europeans. Our transcriptomic results demonstrate that inversions have a direct effect on gene expression patterns not only in inverted regions, but also elsewhere in the genome that could be explained by the disruption of TADs and NHCCs. Moreover, the fact that we have identified several lncRNAs indicates that gene regulation could have an important role in ASD, particularly implicating *KANSL1-AS1* and *LINC00938*. We note that ASD carriers of the susceptibility alleles (inv17q21.31/inv8p23.1), have differential expression of neuronal function-related genes that could be novel ASD candidate genes including *ATP13A1, NR4A3* and *SLC12A6*. We also highlight the implication of a deregulation of immune-inflammatory pathways and actin cytoskeletal regulation in ASD pathogenesis.

Funding

This work was funded by grants from the Spanish Ministry of Economy and Competiveness (FIS PI17/00369). MC-R had a predoctoral fellowship of Ministry of Education, Culture and Sport (FPU16/06907).

Author's contributions

MC-R processed the samples from ASD patients and controls, genotyped the inversions, performed the analysis of RNA sequencing data and drafted the manuscript. ML-S and CA-R acquired the GWAS and SNP array data and carried out the TDT study. TV, PM-C, BG, EG-V and NC recruited the ASD patients included in the transcriptomic study. LAP-J and IC conceived the study and participated in the design and data interpretation. IC helped in drafting the manuscript.

REFERENCES

1. Lord C, Brugha TS, Charman T, Cusack J, Dumas G, Frazier T, et al. Autism spectrum disorder. Nat Rev Dis Prim. 2020;6(1):5.
2. Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The Changing Epidemiology of Autism Spectrum Disorders. Annu Rev Public Health. 2017;38:81–102.
3. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. J Am Acad Child Adolesc Psychiatry. 2017;56(6):466–74.
4. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The Heritability of Autism Spectrum Disorder. JAMA. 2017;318(12):1182–4.
5. Yip BHK, Bai D, Mahjani B, Klei L, Pawitan Y, Hultman CM, et al. Heritable Variation, With Little or No Maternal Effect, Accounts for Recurrence Risk to Autism Spectrum Disorder in Sweden. Biol Psychiatry. 2018;83(7):589–97.
6. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron. 2015;87(6):1215–33.
7. Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet. 2017;49(4):515–26.
8. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180:1–17.
9. Wiśniowiecka-Kowalnik B, Nowakowska BA. Genetics and epigenetics of autism spectrum disorder—current evidence in the field. J Appl Genet. 2019;60(1):37–47.
10. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409(6846):860–921.
11. Numanagić I, Gökkaya AS, Zhang L, Berger B, Alkan C, Hach F. Fast characterization of segmental duplications in genome assemblies. Bioinformatics. 2018;34(17):i706–14.
12. Emanuel BS, Shaikh TH. Segmental duplications: An "expanding" role in genomic instability and disease. Nat Rev Genet. 2001;2(10):791–800.
13. Bailey JA, Eichler EE. Primate segmental duplications: crucibles of evolution, diversity and disease. Nat Rev Genet. 2006;7(7):552–64.
14. Puig M, Casillas S, Villatoro S, Cáceres M. Human inversions and their functional consequences. Brief Funct Genomics. 2015;14(5):369–79.
15. Cukier HN, Skaar DA, Rayner-Evans MY, Konidari I, Whitehead PL, Jaworski JM, et al. Identification of Chromosome 7 Inversion Breakpoints in an Autistic Family Narrows Candidate Region for Autism Susceptibility. 2009;2(5):258–66.
16. González JR, Cáceres A, Esko T, Cuscó I, Puig M, Esnaola M, et al. A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. Am J Hum Genet. 2014;94(3):361–72.
17. González JR, Ruiz-Arenas C, Cáceres A, Morán I, López-Sánchez M, Alonso L, et al. Polymorphic Inversions Underlie the Shared Genetic Susceptibility of Obesity-Related Diseases. Am J Hum Genet. 2020;106(6):846–58.
18. Ruiz-Arenas C, Cáceres A, López-Sánchez M, Tolosana I, Pérez-Jurado L, González JR. scoreInvHap: Inversion genotyping for genome-wide association studies. PLoS Genet. 2019;15(7):1–16.
19. Pain O, Pocklington AJ, Holmans PA, Bray NJ, O'Brien HE, Hall LS, et al. Novel Insight Into the Etiology of Autism Spectrum Disorder Gained by Integrating Expression Data With Genome-wide Association Statistics. Biol Psychiatry. 2019;86(4):265–73.
20. Antonacci F, Kidd JM, Marques-Bonet T, Ventura M, Siswara P, Jiang Z, et al. Characterization of six human disease-associated inversion polymorphisms. Hum Mol

Genet. 2009;18(14):2555–66.
21. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, et al. A common inversion under selection in Europeans. Nat Genet. 2005;37(2):129–37.
22. de Jong S, Chepelev I, Janson E, Strengman E, van den Berg LH, Veldink JH, et al. Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. BMC Genomics. 2012;13(458):1–6.
23. Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum Mol Genet. 1999;8(4):711–5.
24. Houlden H, Baker M, Morris HR, MacDonald N, Pickering-Brown S, Adamson J, et al. Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype. Neurology. 2001;56(12):1702–6.
25. Skipper L, Wilkes K, Toft M, Baker M, Lincoln S, Hulihan M, et al. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. Am J Hum Genet. 2004;75(4):669–77.
26. Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, Fung HC, et al. The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. Hum Mol Genet. 2005;14(16):2399–404.
27. Sánchez-Juan P, Moreno S, de Rojas I, Hernández I, Valero S, Alegret M, et al. The MAPT H1 Haplotype Is a Risk Factor for Alzheimer's Disease in APOE ε4 Non-carriers. Front Aging Neurosci. 2019;11(327):1–9.
28. Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. Nat Genet. 2006;38(9):1032–7.
29. Salm MPA, Horswell SD, Hutchison CE, Speedy HE, Yang X, Liang L, et al. The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. Genome Res. 2012;22:1144–53.
30. Sugawara H, Harada N, Ida T, Ishida T, Ledbetter DH, Yoshiura KI, et al. Complex low-copy repeats associated with a common polymorphic inversion at human chromosome 8p23. Genomics. 2003;82(2):238–44.
31. Bosch N, Morell M, Ponsa I, Mercader JM, Armengol L, Estivill X. Nucleotide, cytogenetic and expression impact of the human chromosome 8p23.1 inversion polymorphism. PLoS One. 2009;4(12):e8269.
32. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, Clinical, and Population Relevance of 95 Loci for Blood Lipids. Nature. 2010;466(7307):707–13.
33. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013;45(11):1283.
34. Cáceres A, González JR. Following the footprints of polymorphic inversions on SNP data: from detection to association tests. Nucleic Acids Res. 2015;43(8):e53:1-11.
35. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, et al. Association of Systemic Lupus Erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med. 2008;358(9):900–9.
36. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Kawamoto M, et al. Association of the FAM167A-BLK region with systemic sclerosis. Arthritis Rheum. 2010;62(3):890–5.
37. Zhou Y, Li X, Wang G, Li X. Association of FAM167A-BLK rs2736340 Polymorphism with Susceptibility to Autoimmune Diseases: A Meta-Analysis. Immunol Invest. 2016;45(4):336–48.
38. Song GG, Lee YH. Association between BLK polymorphisms and susceptibility to SLE. Z Rheumatol. 2017;76(2):176–82.
39. Namjou B, Ni Y, Harley ITW, Chepelev I, Cobb B, Kottyan LC, et al. The effect of inversion at 8p23 on BLK association with Lupus in Caucasian population. PLoS One. 2014;9(12):1–13.
40. Demirci FY, Wang X, Morris DL, Feingold E, Bernatsky S, Pineau C, et al. Multiple signals

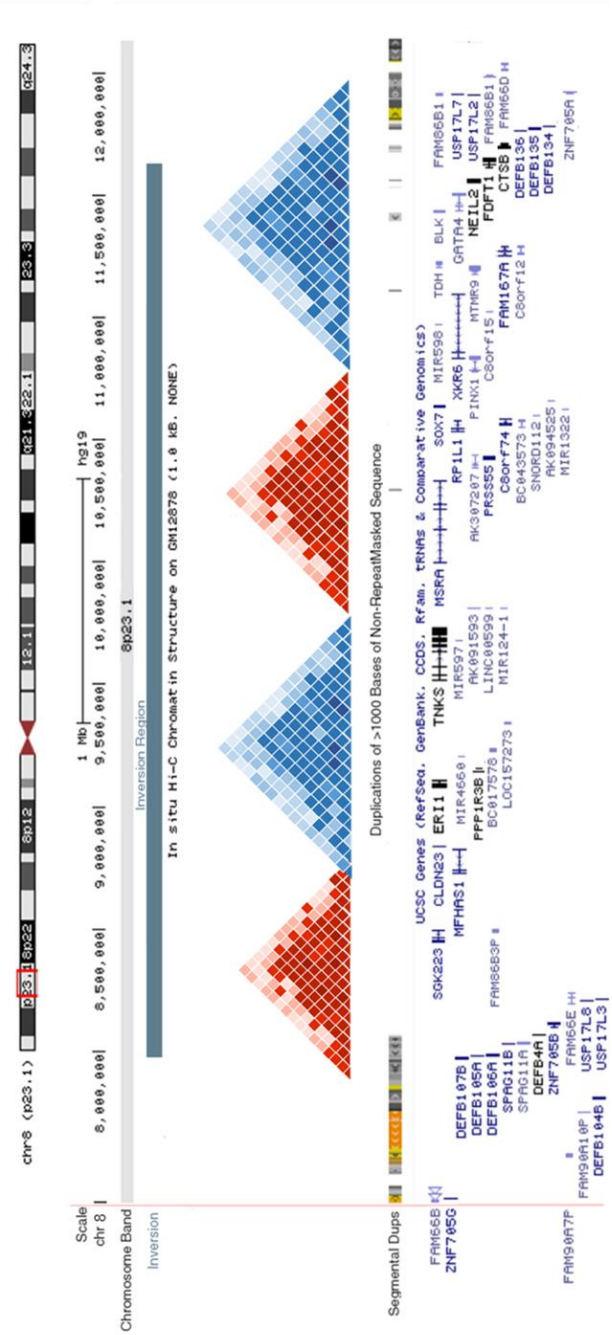
at the extended 8p23 locus are associated with susceptibility to systemic lupus erythematosus (SLE). J Med Genet. 2017;54(6):381–9.
41. Pedersen BS, Quinlan AR. Who's Who? Detecting and Resolving Sample Anomalies in Human DNA Sequencing Studies with Peddy. Am J Hum Genet. 2017;100(3):406–13.
42. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81(3):559–75.
43. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21.
44. Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics. 2015;31(2):166–9.
45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12).
46. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB - A database for integrating human functional interaction networks. Nucleic Acids Res. 2009;37(Database issue):D623-8.
47. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreserv Biobank. 2015;13(5):319.
48. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485(7398):376–80.
49. Liu T, Porter J, Zhao C, Zhu H, Wang N, Sun Z, et al. TADKB: Family classification and a knowledge base of topologically associating domains. BMC Genomics. 2019;20(1):217.
50. Maass PG, Barutcu AR, Rinn JL. Interchromosomal interactions: A genomic love story of kissing chromosomes. J Cell Biol. 2019;218(1):27–38.
51. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet. 2018;19(7):453–67.
52. Kurth I, Klopocki E, Stricker S, Van Oosterwijk J, Vanek S, Altmann J, et al. Duplications of noncoding elements 5′ of SOX9 are associated with brachydactyly-anonychia. Nat Genet. 2009;41:862–3.
53. Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet. 2017;49(1):36–45.
54. Brandler WM, Antaki D, Gujral M, Kleiber ML, Whitney J, Maile MS, et al. Paternally inherited cis-regulatory structural variants are associated with autism. Science. 2018;360(6386):327–31.
55. Rabl C. Über Zelltheilung. Morph Jb. 1885;10:214–330.
56. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. PLoS Biol. 2006;4(5):780–8.
57. Németh A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Péterfia B, et al. Initial genomics of human nucleolus. PLoS Genet. 2010;6(3):e1000889.
58. Kaufmann S, Fuchs C, Gonik M, Khrameeva EE, Mironov AA, Frishman DF. Inter-chromosomal contact networks provide insights into mammalian chromatin organization. PLoS One. 2015;10(5):1–25.
59. Loviglio MN, Leleu M, Männik K, Passeggeri M, Giannuzzi G, Van Der Werf I, et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. Mol Psychiatry. 2017;22(6):836–49.
60. Guttman M, Rinn JL. Modular regulatory principles of large non-coding RNAs. NatureNature. 2014;7385:339–46.
61. Tang J, Yu Y, Yang W. Long noncoding RNA and its contribution to autism spectrum disorders. CNS Neurosci Ther. 2017;23(8):645–56.

62. Ziats MN, Rennert OM. Aberrant Expression of Long Noncoding RNAs in Autistic Brain. J Mol Neurosci. 2013;49(3):589–93.
63. Wang Y, Zhao X, Ju W, Flory M, Zhong J, Jiang S, et al. Genome-wide differential expression of synaptic long noncoding RNAs in autism spectrum disorder. Transl Psychiatry. 2015;5(10).
64. Parikshak NN, Swarup V, Belgard TG, Irimia M, Gandal MJ, Hartl C, et al. Genome-wide changes in lncRNA, splicing and regional gene expression patterns in autism. Nature. 2016;540(7633):423–7.
65. Sayad A, Omrani MD, Fallah H, Taheri M, Ghafouri-Fard S. Aberrant Expression of Long Non-coding RNAs in Peripheral Blood of Autistic Patients. J Mol Neurosci. 2019;67(2):276–81.
66. Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB. A noncoding RNA antisense to moesin at 5p14.1 in autism. Sci Transl Med. 2012;4(128).
67. Gudenas BL, Srivastava AK, Wang L. Integrative genomic analyses for identification and prioritization of long non-coding RNAs associated with autism. PLoS One. 2017;12(5):1–15.
68. Bilinovich SM, Lewis K, Grepo N, Campbell DB. The long noncoding rna rps10p2-as1 is implicated in autism spectrum disorder risk and modulates gene expression in human neuronal progenitor cells. Front Genet. 2019;10(970):1–7.
69. Yoo HJ, Cho IH, Park M, Cho E, Cho SC, Kim BN, et al. Association between PTGS2 polymorphism and autism spectrum disorders in Korean trios. Neurosci Res. 2008;62(1):66–9.
70. Rothwell NJ, Luheshi GN. Interleukin 1 in the brain: Biology, pathology and therapeutic target. Trends Neurosci. 2000;23(12):618–25.
71. Saad K, Abdallah AEM, Abdel-Rahman AA, Al-Atram AA, Abdel-Raheem YF, Gad EF, et al. Polymorphism of interleukin-1β and interleukin-1 receptor antagonist genes in children with autism spectrum disorders. Prog NeuroPsychopharmacology Biol Psychiatry. 2020;103:109999.
72. Pekkoc Uyanik KC, Kalayci Yigin A, Dogangun B, Seven M. Evaluation of IL1B rs1143634 and IL6 rs1800796 Polymorphisms with Autism Spectrum Disorder in the Turkish Children. Immunol Invest. 2021;1–12
73. Papiol S, Rosa A, Gutiérrez B, Martín B, Salgado P, Catalán R, et al. Interleukin-1 cluster is associated with genetic risk for schizophrenia and bipolar disorder. J Med Genet. 2004;41(3):219–23.
74. Zhu X, Wei D, Chen O, Zhang Z, Xue J, Huang S, et al. Upregulation of CCL3/MIP-1alpha regulated by MAPKs and NF-kappaB mediates microglial inflammatory response in LPS-induced brain injury. Acta Neurobiol Exp (Wars). 2016;76(4):304–17.
75. McNair K, Spike R, Guilding C, Prendergast GC, Stone TW, Cobb SR, et al. A role for RhoB in synaptic plasticity and the regulation of neuronal morphology. J Neurosci. 2010;30(9):3508–17.
76. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Wigler M. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216–21.
77. Freed D, Pevsner J. The Contribution of Mosaic Variants to Autism Spectrum Disorder. PLoS Genet. 2016;12(9):1–20.
78. Weingarten LS, Dave H, Li H, Crawford DA. Developmental expression of P5 ATPase mRNA in the mouse. Cell Mol Biol Lett. 2012;17(1):153–70.
79. Anazi S, Maddirevula S, Salpietro V, Asi YT, Alsahli S, Alhashem A, et al. Expanding the genetic heterogeneity of intellectual disability. Hum Genet. 2017;136(11–12):1419–29.
80. Novak G, Zai CC, Mirkhani M, Shaikh S, Vincent JB, Meltzer H, et al. Replicated association of the NR4A3 gene with smoking behaviour in schizophrenia and in bipolar disorder. Genes, Brain Behav. 2010;9(8):910–7.
81. Elmslie F V., Rees M, Williamson MP, Kerr M, Kjeldsen MJ, Pang KA, et al. Genetic mapping of a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q. Hum

Mol Genet. 1997;6(8):1329–34.
82. Moser D, Ekawardhani S, Kumsta R, Palmason H, Bock C, Athanassiadou Z, et al. Functional analysis of a potassium-chloride co-transporter 3 (SLC12A6) promoter polymorphism leading to an additional DNA methylation site. Neuropsychopharmacol. 2009;34(2):458–67.
83. Meyer J, Johannssen K, Freitag CM, Schraut K, Teuber I, Hahner A, et al. Rare variants of the gene encoding the potassium chloride co-transporter 3 are associated with bipolar disorder. Int J Neuropsychopharmacol. 2005;8(4):495–504.
84. Howard HC, Mount DB, Rochefort D, Byun N, Dupré N, Lu J, et al. The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. Nat Genet. 2002;32(3):384–92.
85. Hirabayashi Y, Kwon S, Paek H, Pernice W, Paul M, Lee J, et al. ER-mitochondria tethering by PDZD8 regulates Ca2+ dynamics in mammalian neurons. Science (80-). 2017;358(6363):623–30.
86. Chen Y, Zheng J, Li X, Zhu L, Shao Z, Yan X, et al. Wdr47 Controls Neuronal Polarization through the Camsap Family Microtubule Minus-End-Binding Proteins. Cell Rep. 2020;31(3):107526.
87. Khanzada NS, Butler MG, Manzard AM. Geneanalytics pathway analysis and genetic overlap among autism spectrum disorder, bipolar disorder and Schizophrenia. Int J Mol Sci. 2017;18(3):1–11.
88. Kushima I, Aleksic B, Nakatochi M, Shimamura T, Okada T, Uno Y, et al. Comparative analyses of copy-number variation in Autism spectrum disorder and schizophrenia reveal etiological overlap and biological insights. Cell Rep. 2018;24(11):2838–56.
89. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Mill J, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. 2013;474(7351):380–4.
90. 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):1–8.
91. Hu VW, Frank BC, Heine S, Lee NH, Quackenbush J. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. BMC Genomics. 2006;7:118.
92. Gregg J, Lit L, Baron C, Hertz-Picciotto I, Walker W, Davis R, et al. Gene expression changes in children with autism. Genomics. 2008;91:22–9.
93. Enstrom A, Lit L, Onore C, Gregg J, Hansen R, Pessah I, et al. Altered gene expression and function of peripheral blood natural killer cells in children with autism. Brain Behav Immun. 2009;23(1):124–33.
94. Chien WH, Gau SSF, Chen CH, Tsai WC, Wu YY, Chen PH, et al. Increased gene expression of FOXP1 in patients with autism spectrum disorders. Mol Autism. 2013;4(1):1–8.
95. Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee IH, et al. Characteristics and Predictive Value of Blood Transcriptome Signature in Males with Autism Spectrum Disorders. PLoS One. 2012;7(12):e49475.
96. Pramparo T, Pierce K, Lombardo M, Carter Barnes C, Marinero S, Ahrens-Barbeau C, et al. Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. JAMA Psychiatry. 2015;72(4):386–94.
97. Tylee DS, Hess JL, Quinn TP, Barve R, Huang H, Zhang-James Y, et al. Blood transcriptomic comparison of individuals with and without autism spectrum disorder: A combined-samples mega-analysis. Am J Med Genet Part B Neuropsychiatr Genet. 2017;174(3):181–201.
98. Filosi M, Kam-Thong T, Essioux L, Muglia P, Trabetti E, Spooren W, et al. Transcriptome signatures from discordant sibling pairs reveal changes in peripheral blood immune cell composition in Autism Spectrum Disorder. Transl Psychiatry. 2020;10(1):106.
99. Bjørklund G, Saad K, Chirumbolo S, Kern JK, Geier DA, Geier MR, et al. Immune dysfunction and neuroinflammation in autism spectrum disorder. Acta Neurobiol Exp

(Wars). 2016;76(4):257–68.
100. Siniscalco D, Schultz S, Brigida AL, Antonucci N. Inflammation and Neuro-Immune Dysregulations in Autism Spectrum Disorders. Pharm. 2018;11(56):1–14.
101. Pangrazzi L, Balasco L, Bozzi Y. Molecular Sciences Oxidative Stress and Immune System Dysfunction in Autism Spectrum Disorders. Int J Mol Sci. 2020;21(3288):1–14.
102. Rodriguez JI, Kern JK. Evidence of microglial activation in autism and its possible role in brain underconnectivity. Neuron Glia Biol. 2011;7(2–4):205–13.
103. Voineagu I, Eapen V. Converging pathways in autism spectrum disorders: Interplay between synaptic dysfunction and immune responses. Front Hum Neurosci. 2013;7:738.
104. Sekino Y, Kojima N, Shirao T. Role of actin cytoskeleton in dendritic spine morphogenesis. Neurochem Int. 2007;51(2–4):92–104.
105. Joensuu M, Lanoue V, Hotulainen P. Dendritic spine actin cytoskeleton in autism spectrum disorder. Prog Neuro-Psychopharmacology Biol Psychiatry. 2018;84(Pt B):362–81.
106. Duffney L, Zhong P, Wei J, Matas E, Cheng J, Qin L, et al. Autism-like deficits in Shank3-deficient mice are rescued by targeting actin regulators. Cell Rep. 2015;11(9):1400–13.
107. Griesi-Oliveira K, Suzuki AM, Alves AY, Mafra ACCN, Yamamoto GL, Ezquina S, et al. Actin cytoskeleton dynamics in stem cells from autistic individuals. Sci Rep. 2018;8(1):1–10.

SUPPLEMENTARY MATERIAL



Supplementary Figure 1. Schematic representation of inv8p23.1 and the TADs present in the region

Supplementary Table 1. Characteristics of subjects included in the transcriptomic study.

Patient	SampleID	Phenotype	Genotype inv17q21.31	Genotype inv8p23.1	Sex	RNA extraction
AUT163 01	H5HVGBBXY_2_148UDI	Affected	H1/H1	NI/NI	M	Trizol+Qiagen
AUT198 01	H5HVGBBXY_2_160UDI	Affected	H1/H1	NI/NI	M	Trizol+Qiagen
AUT201 01	H5HVGBBXY_2_172UDI	Affected	H1/H1	NI/NI	M	Trizol+Qiagen
AUT224 01	H5HVGBBXY_2_184UDI	Affected	H1/H1	NI/NI	M	Trizol+Qiagen
AUT266 01	H5HVGBBXY_2_196UDI	Affected	H1/H1	NI/NI	M	Trizol+Qiagen
C33	H5HVGBBXY_2_208UDI	Control	H1/H1	NI/NI	F	Qiagen
29M	H5HVGBBXY_2_220UDI	Control	H1/H1	NI/NI	F	Qiagen
CM1	H5HVGBBXY_2_232UDI	Control	H1/H1	NI/NI	M	Trizol+Qiagen
AUT145 03	H5HVGBBXY_2_155UDI	Affected	H1/H1	I/I	M	Trizol+Qiagen
AUT217 01	H5HVGBBXY_2_167UDI	Affected	H1/H1	I/I	M	Trizol+Qiagen
AUT233 01	H5HVGBBXY_2_179UDI	Affected	H1/H1	I/I	M	Qiagen
AUT280 01	H5HVGBBXY_2_191UDI	Affected	H1/H1	I/I	M	Qiagen
C29	H5HVGBBXY_3_203UDI	Control	H1/H1	I/I	F	Qiagen
C31	H5HVGBBXY_3_228UDI	Control	H1/H1	I/I	F	Qiagen
C38	H5HVGBBXY_3_215UDI	Control	H1/H1	I/I	F	Qiagen
C43	H5HVGBBXY_3_227UDI	Control	H1/H1	I/I	M	Qiagen
AUT143 01	H5HVGBBXY_3_239UDI	Affected	H2/H2	NI/I	M	Trizol+Qiagen
AUT179 01	H5HVGBBXY_3_156UDI	Affected	H2/H2	NI/I	M	Qiagen
AUT301 01	H5HVGBBXY_3_168UDI	Affected	H2/H2	NI/I	M	Qiagen
AUT332 01	H5HVGBBXY_3_180UDI	Affected	H2/H2	NI/I	M	Qiagen
C34	H5HVGBBXY_3_192UDI	Control	H2/H2	NI/NI	F	Qiagen
CM8	H5HVGBBXY_3_204UDI	Control	H2/H2	NI/I	M	Trizol+Qiagen
CM11	H5HVGBBXY_3_216UDI	Control	H2/H2	NI/I	M	Trizol+Qiagen
CF24	H5HVGBBXY_3_240UDI	Control	H2/H2	NI/I	F	Trizol+Qiagen

Supplementary Table 2. TDT analysis in inversions 17q21.31 and 8p23.1.

	Dataset	Selected samples	Total trios	T	NT	%OT	OR	L95	U95	Chisq	p-value
Inv17q21.31	AGP	6395	1868	730	609	9.04	1.20	1.08	1.34	10.930	0.00094
	SSC	5670	1613	572	552	1.78	1.04	0.92	1.17	0.356	0.5508
	UMSGARD	1030	289	108	102	2.86	1.06	0.81	1.39	0.171	0.6788
	MCFTR	5333	1205	408	381	3.42	1.07	0.93	1.23	0.924	0.3364
	IMAGE	2417	686	248	231	3.55	1.07	0.90	1.28	0.603	0.4373
	SHARP	2160	326	86	84	1.18	1.02	0.76	1.38	0.024	0.8781
Inv8p23.1	AGP	6395	1868	955	832	6.88	1.15	1.05	1.26	8.466	0.0036
	SSC	5670	1613	826	757	4.36	1.09	0.99	1.20	3.008	0.0829
	UMSGARD	1030	289	161	145	5.23	1.11	0.89	1.39	0.837	0.3604
	MCFTR	5333	1205	599	569	2.57	1.05	0.94	1.18	0.771	0.3800
	IMAGE	2417	686	303	328	-3.96	0.92	0.79	1.08	0.991	0.3196
	SHARP	2160	326	154	156	-0.64	0.99	0.79	1.23	0.013	0.9096

T: transmitted (I allele), NT: not transmitted, OT: % over-transmission of I allele, OR: odds ratio, L95: lower 95% confidence limit, U95: upper 95% confidence limit.

Supplementary Table 3. TDT categorizing ASD samples according to different subphenotypes.

Inv17q21.31	**Variable**	**Category**	**Trios**	**T**	**NT**	**OR**	**L95**	**U95**	**p-value**
meta-analysis of ASD	family	multiplex	896	268	211	1.27	1.18	1.36	9.20E-03
	IQ	high>80	2089	653	558	1.17	1.11	1.23	6.33E-03
	strict	no	1153	368	284	1.30	1.22	1.37	1.00E-03

Inv8p23.1	**Variable**	**Category**	**Trios**	**T**	**NT**	**OR**	**L95**	**U95**	**p-value**
meta-analysis of ASD	verbal	yes	3123	1050	1270	1.21	1.17	1.25	4.94E-06
	strict	no	1065	341	439	1.29	1.22	1.36	4.50E-04

T: transmitted (I allele), NT: not transmitted, OR: odds ratio, L95: lower 95% confidence limit, U95: upper 95% confidence limit, IQ: Intelligence Quotient.

Supplementary Table 4. Differentially expressed genes between inverted ASD individuals (H2A) and other subjects for inv17q21.31. SE: standard error, stat: Wald statistic.

Ensembl name	Gene name	Gene biotype	Chr band	Base mean	log2FC	lfcSE	stat	p-value	q-value	H2A	Others
ENSG00000201944	*SNORA72*	snoRNA	1q32.1	3.838	-18.799	1.590	-11.824	2.95E-32	8.32E-28	0	4.606
ENSG00000254693	*RP11-58K22.5*	lncRNA	11p11.2	3.582	-18.139	1.741	-10.421	1.98E-25	2.48E-21	0	4.298
ENSG00000225420	*AC104134.2*	lncRNA	2p11.2	3.721	-18.316	1.789	-10.237	1.36E-24	9.56E-21	0	4.465
ENSG00000244080	*RN7SL833P*	pseudogene	19p13.2	0.721	-17.313	1.923	-9.002	2.23E-19	6.98E-16	0	0.865
ENSG00000259589	*RP11-317G6.1*	lncRNA	15q22.31	0.769	-16.951	2.243	-7.558	4.08E-14	8.85E-11	0	0.923
ENSG00000214425	*LRRC37A4P*	pseudogene	17q21.31	243.440	-5.487	0.800	-6.858	7.00E-12	1.10E-08	7.759	290.577
ENSG00000242463	*RN7SL500P*	pseudogene	22q13.33	0.452	-16.384	2.585	-6.339	2.31E-10	3.43E-07	0	0.542
ENSG00000214401	*KANSL1-AS1*	lncRNA	17q21.31	79.156	2.939	0.515	5.705	1.16E-08	1.56E-05	238.227	47.342
ENSG00000073756	*PTGS2*	protein coding	1q31.1	1070.719	3.787	0.834	4.544	5.53E-06	0.0057	1191.697	1046.523
ENSG00000125538	*IL1B*	protein coding	2q14.1	1291.616	3.493	0.769	4.539	5.65E-06	0.0057	925.363	1364.866
ENSG00000163739	*CXCL1*	protein coding	4q13.3	123.677	5.533	1.242	4.456	8.36E-06	0.0081	245.143	99.384
ENSG00000143878	*RHOB*	protein coding	2p24.1	1671.689	2.391	0.543	4.405	1.06E-05	0.0099	5278.956	950.236
ENSG00000255092	*RP11-58K22.4*	lncRNA	11p11.2	34.688	-5.246	1.192	-4.400	1.08E-05	0.0099	0.486	41.528
ENSG00000204792	*LINC01291*	lncRNA	2p12	59.933	-7.155	1.681	-4.257	2.07E-05	0.0182	1.022	71.715
ENSG00000006075	*CCL3*	protein coding	17q12	114.233	4.313	1.019	4.234	2.30E-05	0.0197	169.801	103.120
ENSG00000163492	*CCDC141*	protein coding	2q31.2	82.421	-1.315	0.315	-4.178	2.95E-05	0.0245	33.874	92.131
ENSG00000120949	*TNFRSF8*	protein coding	1p36.22	208.947	1.500	0.364	4.123	3.73E-05	0.0301	457.218	159.293
ENSG00000233264	*AC006042.8*	pseudogene	7p21.3	64.642	-1.922	0.470	-4.092	4.28E-05	0.0335	18.854	73.800
ENSG00000273015	*LINC00938*	lncRNA	12q12	422.595	-0.559	0.139	-4.009	6.10E-05	0.0465	320.276	443.058

Supplementary Table 5. Differentially expressed genes between inverted ASD individuals (IA) and other subjects for inv8p23.1. SE: standard error, stat: Wald statistic.

Ensembl name	Gene name	Gene biotype	Chr band	Base mean	log2FC	lfcSE	stat	p-value	q-value	IA	Others
ENSG00000119508	*NR4A3*	protein coding	9q22	285.062	4.986	0.943	5.288	1.24E-07	0.001	716.061	198.863
ENSG00000198336	*MYL4*	protein coding	17q21.32	68.262	5.186	0.992	5.228	1.71E-07	0.001	348.097	216.097
ENSG00000105726	*ATP13A1*	protein coding	19p13.11	1589.772	-0.570	0.118	-4.845	1.27E-06	0.003	1203.018	1667.123
ENSG00000232891	*RP11-136K14.1*	lncRNA	6q25.1	12.877	4.600	0.942	4.880	1.06E-06	0.003	24.890	10.474
ENSG00000237719	*RP11-179N16.3*	pseudogene	6q21.31	81.030	2.108	0.435	4.846	1.26E-06	0.003	164.353	64.365
ENSG00000244071	*RPL9P33*	pseudogene	19q13.41	25.546	3.132	0.675	4.640	3.48E-06	0.007	76.099	15.435
ENSG00000110917	*MLEC*	protein coding	12q24.31	1200.148	-0.874	0.192	-4.560	5.11E-06	0.008	747.823	1290.613
ENSG00000267758	*RP11-358B23.5*	lncRNA	17q21.2	23.350	1.798	0.393	4.569	4.89E-06	0.008	48.133	18.394
ENSG00000232284	*GNG12-AS1*	lncRNA	1p31.3	6.190	2.492	0.551	4.518	6.23E-06	0.009	19.730	3.482
ENSG00000085433	*WDR47*	protein coding	1p13.3	365.585	0.748	0.171	4.368	1.25E-05	0.016	528.561	332.990
ENSG00000133026	*MYH10*	protein coding	17p13.1	229.434	1.570	0.358	4.381	1.18E-05	0.016	495.648	176.191
ENSG00000236194	*AC003104.1*	lncRNA	17q21.2	39.176	1.817	0.417	4.361	1.29E-05	0.016	80.574	30.897
ENSG00000158856	*DMTN*	protein coding	8p21.3	466.260	1.921	0.455	4.224	2.40E-05	0.027	1060.990	1227.692
ENSG00000205571	*SMN2*	protein coding	5q13.2	52.038	-3.291	0.781	-4.213	2.52E-05	0.027	6.252	61.195
ENSG00000229598	*PRDX3P1*	pseudogene	22q13.1	109.988	2.586	0.621	4.168	3.08E-05	0.030	199.724	92.041
ENSG00000243650	*RN7SL834P*	pseudogene	2q37.1	61.899	1.419	0.340	4.168	3.08E-05	0.030	104.898	53.300
ENSG00000073969	*NSF*	protein coding	17q21.31	586.826	-0.694	0.169	-4.117	3.84E-05	0.035	398.143	624.563
ENSG00000101811	*CSTF2*	protein coding	Xq22	189.213	-0.784	0.191	-4.099	4.16E-05	0.036	121.772	202.701
ENSG00000233175	*CTD-2020K17.3*	lncRNA	17q21.31	86.974	1.411	0.350	4.028	5.63E-05	0.047	166.824	71.004

(Continued)

Ensembl name	Gene name	Gene biotype	Chr band	Base mean	log2FC	lfcSE	stat	p-value	q-value	IA	Others
ENSG00000112486	*CCR6*	protein coding	6q27	24.524	1.532	0.389	3.939	8.17E-05	0.049	47.732	19.882
ENSG00000136732	*GYPC*	protein coding	2q14.3	2137.878	1.974	0.500	3.949	7.85E-05	0.049	6093.484	4537.553
ENSG00000145945	*FAM50B*	protein coding	6p25.2	57.356	-1.058	0.265	-4.000	6.34E-05	0.049	31.939	62.439
ENSG00000154319	*FAM167A*	protein coding	8p23.1	58.321	-1.817	0.462	-3.933	8.39E-05	0.049	17.738	66.438
ENSG00000154328	*NEIL2*	protein coding	8p23.1	180.553	-0.602	0.154	-3.899	9.65E-05	0.049	134.006	189.862
ENSG00000159079	*C21orf59*	protein coding	21q22.11	76.135	-0.982	0.249	-3.947	7.92E-05	0.049	48.487	81.665
ENSG00000165650	*PDZD8*	protein coding	10q25.3-q26.11	1242.155	1.037	0.261	3.972	7.12E-05	0.049	2021.587	1086.269
ENSG00000228363	*AC015971.2*	lncRNA	2p11.2	194.502	1.726	0.439	3.928	8.56E-05	0.049	350.868	163.229
ENSG00000252391	*RNU6-638P*	pseudogene	17q25.3	21.987	2.312	0.590	3.919	8.90E-05	0.049	29.457	20.493
ENSG00000258667	*HIF1A-AS2*	lncRNA	14q23.2	16.907	2.642	0.676	3.906	9.37E-05	0.049	34.446	13.399
ENSG00000259118	*RP11-840I19.3*	lncRNA	14q23.3	87.105	1.737	0.446	3.894	9.85E-05	0.049	164.773	71.571
ENSG00000262944	*RP11-473I1.9*	lncRNA	16p13.2	1005.342	1.425	0.357	3.988	6.65E-05	0.049	1784.877	849.435
ENSG00000266320	*MIR3909*	miRNA	22q12.3	7.881	2.650	0.678	3.909	9.27E-05	0.049	19.835	5.490
ENSG00000268852	*AC132872.2*	protein coding	17q25.3	242.050	2.029	0.517	3.924	8.72E-05	0.049	408.049	208.851
ENSG00000271933	*RP11-338I21.1*	lncRNA	10q23.1	85.234	1.580	0.398	3.974	7.07E-05	0.049	145.858	73.109
ENSG00000140199	*SLC12A6*	protein coding	15q14	2535.964	1.140	0.294	3.880	0.000105	0.049	4008.910	2241.374
ENSG00000152359	*POC5*	protein coding	5q13.3	219.198	-0.687	0.177	-3.881	0.000104	0.049	158.215	231.395

CHAPTER 4

Identification of recurrently mutated genes and commonly deregulated pathways in Autism Spectrum Disorder multiple-hit model

Mar Costa-Roger[1], Marta Codina-Solà[1,2], Marcos López-Sánchez[1], Luis A. Pérez-Jurado[1,3], Ivon Cuscó[1,2]

In preparation

ABSTRACT

Background. Autism Spectrum Disorder (ASD) is a heterogeneous group of neurodevelopmental disorders characterised by alterations in the social and communication areas as well as restricted interests and behaviours. Its etiology is heterogeneous and complex, including polygenic models of multiple loci that implicate the contribution of various rare genetic variants. Our objective is to identify the presence of genes that recurrently accumulate rare genetic variants in ASD individuals and to elucidate common deregulated pathways.

Methods. We performed a meta-analysis of exomic data from 4 different cohorts downloaded from publicly available servers (dbGAP and NIMH Data Archive) with a total sample size of 1,078 ASD patients and 747 control individuals. To identify rare genetic variants, we filtered missense, nonsense and splicing variants with a frequency lower than 1/500 according to ExAC database. To elucidate deregulated pathways, we performed a pathway enrichment analysis using the hypergeometric test from KEGG.db package. We selected both genes and pathways found in 10 or more ASD individuals and 0-1 controls with a p-value<0.05.

Results. We identified a total of 44 genes that recurrently accumulate mutations in the ASD population (p-value<0.05). In particular, we found *LHX1* gene exclusively mutated in ASD patients and statistically significant after multiple sample correction (q-value=0.0292). Moreover, other interesting candidates have been identified due to previous associations with ASD and comorbid features or relevant functions in neurons. The accumulation of rare genetic variants in ASD patients showed the specific deregulation of Rheumatoid arthritis pathway (p-value=0.0138).

Conclusions. The identification of regions that accumulate rare genetic variants reinforces the implication of multiple-hit models in the ASD etiology. We have discovered several genes recurrently accumulating rare genetic variants in our ASD cohort, which could contribute to ASD risk in a multiple-hit model highlighting *LHX1, PRRG4* and *PTPN1*. In addition, we have identified Rheumatoid Arthritis signaling pathway as significantly enriched in ASD probands, which further supports a common genetic background between ASD and Rheumatoid Arthritis.

Keywords: ASD, rare variants, pathways, multiple-hit

[1]Genetics Unit, Universitat Pompeu Fabra, Hospital del Mar Research Institute (IMIM), Parc de Salut Mar and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Universitat Pompeu Fabra, Departament de Ciències Experimentals i de la Salut, Barcelona, Spain

[2]Department of Clinical and Molecular Genetics and Medicine Genetics Group, VHIR, University Hospital Vall d'Hebron, Barcelona, Spain

[3]Genetics Service, Parc de Salut Mar, Barcelona, Spain

INTRODUCTION

Autism Spectrum Disorder (ASD) is a heterogeneous group of neurodevelopmental conditions characterised by impairments in the social and communication areas together with restricted, repetitive interests and behaviours (1,2). Its estimated prevalence in developed countries is around 1.5% becoming a major health issue and males are more likely to be diagnosed of ASD than females with a 3-4:1 ratio (3,4).

Family and twin studies have shown a strong genetic involvement in this disorder and ASD heritability has been estimated to be around 83-85% (5,6). Many studies have focused on unravelling the genetic causes underlying this condition, but the ASD etiology is still unclear in many cases. Up to date, it is possible to recognize the genetic etiology in only ~25-35% of patients diagnosed with ASD. The current identified genetic causes modelling ASD genetic liability include single gene disorders associated with autistic features, chromosomal abnormalities, copy number variation (CNV, microduplications and microdeletions), single nucleotide variants (SNVs) and common risk alleles from SNPs (Single Nucleotide Polymorphisms) (7,8). Therefore, ASD genetic liability is shaped by different variants including rare or *de novo* SNVs and CNVs with moderate or high penetrance as well as common variants (SNPs) accounting for a small effect size (9).

Regarding the inheritance model, ASD are considered complex since different patterns of inheritance have been described including monogenic models and more complex forms such as polygenic (or multiple-hit) and multifactorial models. Monogenic models are explained by a single highly penetrant gene mutation or CNV as the major contributor to ASD. In polygenic models, the addition of multiple variants, including both rare and common, accounts for the phenotype of ASD (7,8). In multifactorial models, genotype and environment contribute to the genetic landscape of ASD (9,10).

According to polygenic models, the ASD phenotype may appear when the accumulation of risk variants in certain genes exceeds a theoretical risk threshold (11). The advent of next-generation sequencing (NGS) techniques has drastically accelerated gene discovery in ASD highlighting the large number of genes (>1000) implicated (12,13). A recent study identified 69 high-confidence ASD risk genes with a False Discovery Rate (FDR)<0.1 using a TADA (Transmission And *De novo* Association) approach which integrates *de novo*, transmitted and case-control variation (14). Some examples are *CHD8*, *SCN2A*, *ARID1B, SYNGAP1, NRXN1, SHANK3, PTEN* and *DYRK1A,* among others. ASD risk genes can be grouped in two broad biological classes: synaptic function and transcription/chromatin remodelling (15,16). In addition, several disrupted signaling pathways have been described including glutamatergic synapse, WNT/β-catenin signaling, PI3K/AKT/mTOR pathway, RAS/ERK pathway and MAPK signaling pathway (17–19).

In the present work, we have performed a meta-analysis of exomic data from four different cohorts downloaded from publicly available servers and carried out a genome-wide unbiased search for rare genetic variants following the strategy depicted in Figure 1. The aim of this study is to determine the genetic contribution in multiple-hit models through the identification of susceptibility genes as well as functional pathways overrepresented due to the presence of rare variants in ASD patients.

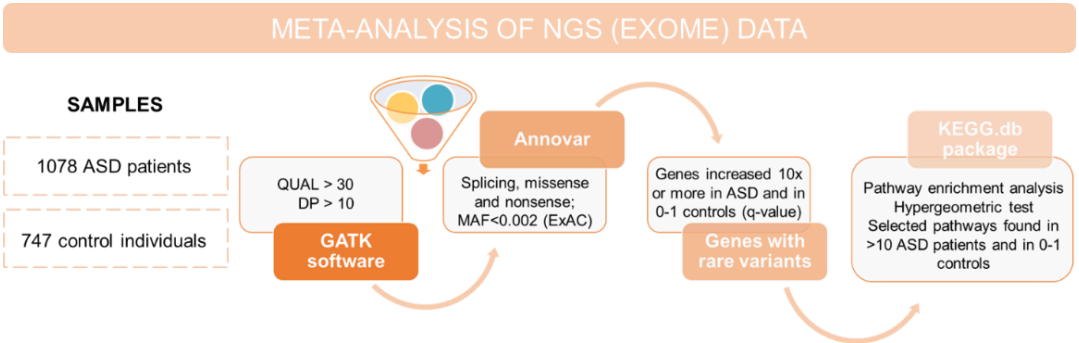


Figure 1. Workflow of the meta-analysis of exome data. NGS: Next Generation Sequencing, ASD: Autism Spectrum Disorder, QUAL: Read quality, DP: Depth of coverage, GATK: Genome Analysis Tool Kit, MAF: Minor Allele Frequency, ExAC: Exome Aggregation Consortium, KEGG: Kyoto Encyclopedia of Genes and Genomes.

METHODS

Study populations

We used exome data from four different datasets downloaded from publicly available repositories including the NCBI database of genotypes and phenotypes (dbGaP; study accession numbers phs000482.v1.p1, phs000298.v3.p2 and phs000616.v2.p2) and the National Institute of Mental Health (NIMH) Data Archive (doi: 10.1038/nature10945) (20). The total sample size was 1,078 ASD patients and 747 control individuals.

WES (Whole Exome Sequencing) samples from the phs000482.v1.p1 study (21) corresponded to 189 cases of sporadic autism. Exonic sequences were enriched through hybridization with a NimbleGen EZ Exome V2.0 and sequencing was performed using an Illumina GAIIx or HiSeq 2000. Regarding the phs000298.v3.p2 study (ARRA Autism Sequencing Collaboration), we used data from 651 autistic patients and 82 controls. Exome capture was carried out using an Applied Biosystems enrichment kit and sequencing by a SOLiD platform. For the phs000616.v2.p2 study (22), we downloaded data from 451 control samples from the Viva La Familia study. The exonic enrichment kit used was NimbleGen VCRome 2.1 (HGSC design) and sequencing was performed using a HiSeq 2000. The last study from the NIMH (10.1038/nature10945) included 238 autism probands and 214 non-affected siblings, which we considered as controls in our meta-analysis. Exonic regions were enriched using NimbleGen custom WE array or EZExomeV2.0 and the captured DNA was sequenced using an Illumina GAIIx or HiSeq 2000. All this information is summarised in Table 1.

Whole-exome sequencing analysis

From the datasets provided by dbGaP (phs000482.v1.p1, phs000298.v3.p2 and phs000616.v2.p2), we downloaded files in SRA format, which were converted to SAM files using the SRA Toolkit command 'sam-dump' and subsequently, BAM files were obtained with samtools. At last, we used 'HaplotypeCaller' from GATK software (23) to generate variant calling files (VCF). Exomes from NIMH Data Archive, were downloaded in VCF format.

Once we obtained all VCF files, we applied quality filters using GATK software and we selected variants with a read quality above 30 (QUAL > 30) and a depth of coverage of at least 10x (DP > 10). Annotation of variants was performed using ANNOVAR software (24), including gene-based and filter-based annotations.

Variant prioritization

In order to identify rare genetic variants that could contribute in a multiple-hit model of ASD, we selected missense, nonsense and splicing variants with a minor allele frequency (MAF) lower than 0.002 according to ExAC database (25). Considering the multiple-hit model, the frequency of each variant should be rare and therefore, we set our MAF threshold in 1/500. Differences regarding ethnicities between datasets were taken into account by selecting specific population frequencies. Non-finish European (NFE) MAF was used for both phs000482.v1.p1 and 10.1038/nature10945 datasets, while for controls from phs000616.v2.p2 we based on Latino/Admixed American (AMR) MAF. As individuals from the phs000298.v3.p2 dataset were of mixed origins, we decided to use the general MAF for all populations.

Enrichment analysis

To identify which genes are recurrently mutated, from the list of rare detected variants we selected those genes that harbored rare genetic variants in 10 or more ASD individuals, but in a very low frequency in the control group (0-1 controls). Chi-squared test was used to obtain p-values. We assessed statistical significance using "qvalue" package (26) from Bioconductor in order to perform a statistical correction for multiple comparisons. We only kept genes with rare genetic variants in more than one dataset in order to minimize possible capture biases between analyzed datasets. In addition, we manually curated all rare genetic variants from the selected genes, excluding those with RF (random forest) quality flag according to gnomAD (27), which indicates variants that are likely to be false-positive calls.

To elucidate common ASD deregulated pathways affected by rare genetic variants, we performed a pathway enrichment analysis using the hypergeometric test from KEGG.db package (28). To identify these deregulated pathways, we performed the enrichment pathway analysis individually including all genes harboring a rare genetic variant in each sample.

The analysis was carried out both in ASD cases and controls in order to determine common ASD deregulated pathways removing those found recurrently altered in controls. This is why we selected pathways found in more than 10 ASD patients and in 0 or 1 controls and with a p-value<0.05 in order to shed light on the biological pathways underlying ASD.

Dataset	Samples	Capture kit	Sequencing platform
phs000482.v1.p1	189 sporadic autism	NimbleGen EZ Exome V2.0	Illumina GAIIx or HiSeq 2000
phs000298.v3.p2	651 autistic patients and 82 controls	Applied Biosystems	SOLiD
phs000616.v2.p2	451 controls	NimbleGen VCRome 2.1 (HGSC design)	HiSeq 2000
10.1038/nature10945 (doi)	238 autism probands and 214 non-affected siblings (controls)	NimbleGen custom WE array or EZExomeV2.0	Illumina GAIIx or HiSeq 2000

Table 1. Summarized information of the exome datasets used for the meta-analysis. Including characteristics regarding the capture kit and sequencing platform as well as the number of samples.

RESULTS

Genes enriched with rare genetic variants

To identify genes that could contribute in a polygenic model, we analyzed exome data from 1,078 ASD patients and 747 control individuals. After quality filters (QUAL>30 and DP>10), we selected splicing, missense and nonsense variants with a MAF<0.05 according to ExAC database. We selected genes that accumulated rare variants in at least 10 ASD patients, but were absent or only found in one control individual.

We found 44 genes recurrently mutated in the ASD subjects (p-value<0.05) and that were present in more than one dataset (Supplementary Table 1), from which five were statistically significant when assessing statistical significance through q-value estimation for FDR control (q-value<0.05). Interestingly, *LHX1* appears exclusively mutated in the ASD cohort (see figure 2). In addition, from the initial 44 genes, 13 were exclusively mutated in the ASD cohort without significance for q-value, such as *DUSP6*, which has been previously associated with bipolar disorder.

From the 44 identified genes, 13 (29.55%) were highly expressed in the brain and 7 (15.91%) have been previously related to ASD or another neurologic disorder such as bipolar disorder, Parkinson's disease or Alzheimer's disease. In particular, 3 genes have been associated with ASD including *LHX1, PRRG4* and *PTPN1*. In addition, 5 (11.36%) genes had brain-related functions highlighting *CARM1, FOXB1* and *CDK5R2*.

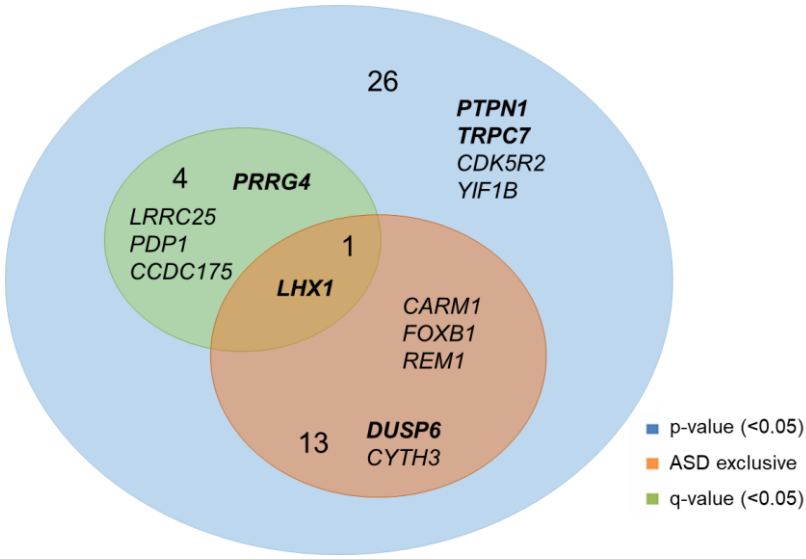


Figure 2. Genes harboring rare variants. Genes in the blue area include 44 genes recurrently mutated in the ASD population, with p-value<0.05 and present in more than one dataset. Genes in orange are ASD exclusive, whereas green genes are statistically significant after statistical correction for multiple comparisons (q-value<0.05). Genes highlighted in bold are genes previously associated to ASD or bipolar disorder.

Deregulated pathways due to the accumulation of genes with rare genetic variants

The accumulation of rare genetic variants in certain ASD risk genes only retrieved the Rheumatoid arthritis pathway as significantly enriched in ASD patients (p-value=0.0138) after performing the pathway enrichment analysis. See table 2. Gap junction and RNA transport were statistically significant ASD-enriched pathways (p-value=0.0022 and p-value=0.0387, respectively), but were also present in two or more controls.

Pathway	ASD patients	Controls	p-value
Rheumatoid arthritis	11	0	0.0138
RNA transport	14	2	0.0386
GAP junction	30	5	0.0022

Table 2. Statistically significant deregulated pathways identified in the enrichment analysis. The total number of individuals with these pathways enriched due to the accumulation of rare genetic variants is indicated as well as the p-value.

DISCUSSION

The use of NGS data, including WES and WGS (Whole Genome Sequencing), has demonstrated to be a powerful approach in the identification of novel ASD candidate genes as well as common deregulated pathways (29,30). In polygenic models, multiple risk variants in certain genes result in exceeding a theoretical risk threshold from which the ASD phenotype appears (11). Therefore, the identification of genes harboring rare genetic variants could contribute to shed light to the ASD etiology.

We have performed a genome-wide search of genes with rare genetic variants identifying 44 genes significantly enriched in ASD population and with p-value<0.05. From this set of genes, the most remarkable is *LHX1*, which is exclusively mutated in 17 ASD patients and statistically significant after multiple testing correction (q-value=0.0292). *LHX1* (LIM homeobox 1) encodes a transcription factor from a large protein family containing the LIM domain and it is expressed in the brain. Evidence suggests that this gene might be involved

in normal brain development as it has been linked to Purkinje cell development in the cerebellum as well as has been identified as a regulator of vertebrate head organization and motor axon guidance to the limbs (31–33). *LHX1* is located in the chromosome 17, concretely in 17q12, a hotspot for genomic disorders where recurrent CNVs are reported. It has been proposed as a candidate gene for the neurological phenotype observed in 17q12 duplications and deletions which both include autism as part of its clinical features (34,35).

Additionally, two more genes were previously associated with ASD (*PRRG4* and *PTPN1*), which are interesting candidates in our analysis. Proline rich and Gla domain 4 (*PRRG4*) was statistically significant after multiple testing correction (q-value=2.12E-04) and was altered in 35 ASD patients compared to only 1 control. This gene has gained some attention as it has been proposed as a functional homologue of the *Drosophila commissureless* gene implicated in axon guidance (36). These findings lead to suggest *PRRG4* as the responsible for the autistic phenotype of the WAGR (Wilm's Tumor, Aniridia, Genitourinary anomalies and Mental Retardation) syndrome (OMIM #194072) caused by interstitial deletions of 11p13 (36,37). On the other hand, *PTPN1* (Protein tyrosine phosphatase non-receptor type 1) gene encodes the PTP1B protein, which is a direct target of MECP2, and was mutated in 16 ASD patients compared to 1 control. A study found that *Mecp2*-mutant mice had elevated PTP1B protein levels indicating that it could be a specific marker of Rett syndrome and therefore an ideal therapeutic target as its inhibition ameliorated neural and behavioral symptoms in RTT mice (38). Moreover, PTP1B activation was related to autism-like behaviours in a mouse model lacking *Lmo4* in interneurons (PV-*Lmo4*KO) (39).

We also highlight several genes that have been linked to bipolar disorder (*DUSP6* and *TRPC7*) or with brain-related functions (*CARM1*, *FOXB1* and *CDK5R2*). Dual specificity phosphatase 6 (*DUSP6*) plays an important role in the negative regulation of MAP kinase superfamily with specificity for the extracellular signal-regulated kinase (ERK) family (40). *DUSP6* was exclusively mutated in our ASD cohort almost reaching significance at FDR level (q-

value=0.0994) and it is located in 12q22-23 region, which is a candidate susceptibility locus for bipolar disorder. In fact, several SNPs in this gene showed positive allelic association in women with bipolar disorder suggesting a possible role in the development of this disorder in women (41,42). *TRPC7* (Transient receptor potential cation channel subfamily C member 7) encodes a transmembrane protein highly expressed in the brain with a possible role as a Ca^{2+} channel. A study using *TRPC7* KO mice discovered that this gene is critical for the generation of seizures both *in vivo* and *in vitro* as it is involved in spontaneous epileptiform bursting in CA3 pyramidal neurons (43). This gene has also been implicated in the abnormal Ca^{2+} homeostasis exhibited in bipolar disorder patients (44). In addition, *TRPC7* is a highly constrained gene (z=2.31) with all variants detected in our analysis being missense, suggesting a possible alteration of the gene. Interestingly, ASD has been reported to increase the risk of major psychiatric disorders, and concretely, a recent study identified that 5% of ASD patients have co-occurring bipolar disorders (45). There is multiple evidence that *CARM1* (Coactivator associated arginine methyltransferase 1), a gene exclusively mutated in 11 ASD patients, has regulatory functions in neurons including negative regulation of synaptic genes and inhibition of neuronal differentiation. A recent study found that *CARM1* inhibition potentiates dendritic complexity and arborization as well as spine maturation during neuronal differentiation (46). Forkhead box B1 (*FOXB1*) is a transcription factor expressed by neural progenitor cells (NPCs) from the embryonic neuroepithelium and it has been proposed as a neuronal modulator in oligodendrocyte development (47). Lastly, *CDK5R2* (Cyclin dependent kinase 5 regulatory subunit 2) encodes a neuron-specific activator of CDK5 kinase and it is essential for oligodendrocyte maturation and myelination as well as dendritic morphogenesis (48,49). More attention should be drawn to these genes in the ASD research as might be new susceptibility genes for this disorder.

Regarding the identification of deregulated pathways, we performed pathway enrichment analysis individually including all genes that harbored rare genetic variants in each subject in order to detect recurrent pathways in ASD patients compared to controls. Only the Rheumatoid arthritis pathway was found

significantly enriched in the ASD population (p-value=0.0138). Multiple evidence has reported an association between family history of autoimmune diseases and ASD, but no comorbidity of Rheumatoid arthritis (RA) in ASD patients has been described so far. Specifically, for maternal RA, researchers found controversial results. First studies reported that 46% of ASD patients had relatives with juvenile or adult RA compared to 26% in controls (OR=2.4, p-value=0.038) and that an increased risk of ASD was observed among children with maternal family history of RA (50,51). Later on, these previous findings were confirmed by several studies where maternal RA was associated with a 30-51% higher risk of ASD in offspring (52–54). It is hypothesized that maternal autoimmunity could negatively affect fetal brain development inducing MAR (Maternal Autoantibody Related) autism (55). Controversially, a study did not find a significant association between women with RA and increased ASD in offspring, although the reported relative risk was 1.42 (95%CI 0.60-3.40) (56) and a meta-analysis showed no substantial association between maternal RA and ASD risk in offspring (OR=0.71, p-value=0.695) (57). Also, it has been reported that family history of non-autoimmune diseases, including autism, was related to an increased risk of RA (58). Furthermore, common genetic predisposition has been observed in individuals with ASD and RA, especially in allelic products of certain genes of the major histocompatibility complex (MHC). The null allele of the *C4B* gene, the extended haplotype B44-SC30-DR4, HLA-A2 and the third hypervariable region of HLA-DRβ1 (concretely HLA-DR4) not only have been found with a significantly higher frequency in ASD individuals and their mothers than controls, but also have been identified as susceptibility markers for RA (59–63). This data suggests that ASD and RA may share common genetic pathways, which could be explained by a model of pleiotropic effects from risk alleles (64).

Limitations

As datasets were different regarding capture kit and sequencing, we tried to avoid possible biases by only selecting those genes with rare genetic variants present in more than one dataset. Besides, in order to overcome ethnical differences between datasets we selected specific population frequencies.

Lastly, frameshift variants were not annotated in the VCF file from NIMH Data Archive (10.1038/nature10945) and therefore, this type of variants have not been evaluated.

Conclusions

In summary, we have discovered several genes that recurrently accumulate rare genetic variants in the studied ASD cohort. As particularly compelling candidates there are *LHX1, PRRG4* and *PTPN1,* which have been previously linked to ASD, thus we provide new evidence to support its relation with this disorder. The remaining highlighted genes could be proposed as new ASD susceptibility genes under a multiple-hit model for its implication in ASD-comorbid conditions such as bipolar disorder (*DUSP6* and *TRPC7*) or its biological relevance in central nervous system (*CARM1, FOX1B* and *CDK5R2*). Lastly, we have identified Rheumatoid Arthritis signaling pathway as significantly enriched in ASD probands due to the accumulation of rare genetic variants, which further supports a common genetic background between ASD and RA.

Abbreviations

ASD: Autism Spectrum Disorder, CNV: copy number variant, SNV: single nucleotide variant, SNPs: single nucleotide polymorphisms, NGS: next generation sequencing, FDR: False Discovery Rate, TADA: Transmission And De novo Association, QUAL: read quality, DP: depth of coverage, GATK: Genome Analysis Tool Kit, MAF: Minor Allele Frequency, NFE: non-Finish European, AMR: Latino/admixed American, ExAC: Exome Aggregation Consortium, KEGG: Kyoto Encyclopedia of Genes and Genomes, dbGaP: data base of Genotypes and Phenotypes, NIMH: National Institute of Mental Health, WES: whole exome sequencing, VCF: variant calling file, WGS: whole genome sequencing, RA: Rheumatoid arthritis.

Ethical approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The datasets are available at dbGaP and NIMH Data Archive.

Competing interests

LAP-J is scientist advisor of qGenomics SL. The remaining authors declare no conflict of interest.

Funding

This work was supported by FIS PI16/00369 grant from the Spanish Ministry of Economy and Competiveness. MC-R had a predoctoral fellowship of Ministry of Education, Culture and Sport (FPU16/06907).

Authors' contributions

MC-R and MC-S analysed exome data and performed enrichment analysis. MC-S and ML-S carried out data acquisition. ML-S provided bioinformatics advice. MC-R drafted the manuscript. IC and LAP-J conceived the study and participated in the design and data interpretation, and helped in drafting the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fifth edition. 5th ed. American Psychiatric Association, editor. Arlington, VA; 2013.
2. Lai MC, Lombardo M V., Baron-Cohen S. Autism. Lancet. 2014;383(9920):896–910.
3. Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The Changing Epidemiology of Autism Spectrum Disorders. Annu Rev Public Health. 2017;38:81–102.
4. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. J Am Acad Child Adolesc Psychiatry. 2017;56(6):466–74.
5. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The Heritability of Autism Spectrum Disorder. JAMA. 2017;318(12):1182–4.
6. Yip BHK, Bai D, Mahjani B, Klei L, Pawitan Y, Hultman CM, et al. Heritable Variation, With Little or No Maternal Effect, Accounts for Recurrence Risk to Autism Spectrum Disorder in Sweden. Biol Psychiatry. 2018;83(7):589–97.
7. Wiśniowiecka-Kowalnik B, Nowakowska BA. Genetics and epigenetics of autism spectrum disorder—current evidence in the field. J Appl Genet. 2019;60(1):37–47.
8. Toma C. Genetic Variation across Phenotypic Severity of Autism. Trends Genet. 2020;36(4):228–31.
9. Iakoucheva LM, Muotri AR, Sebat J. Getting to the Cores of Autism. Cell. 2019;178(6):1287–98.
10. Chaste P, Roeder K, Devlin B. The Yin and Yang of Autism Genetics: How Rare De Novo and Common Variations Affect Liability. Annu Rev Genom Hum Genet. 2017;18(921):1–9.
11. Ramaswami G, Geschwind DH. Genetics of autism spectrum disorder. 1st ed., Handbook of Clinical Neurology. Elsevier B.V. 2018; 147: 321–329.
12. Butler MG, Rafi SK, Manzardo AM. High-resolution chromosome ideogram

representation of currently recognized genes for Autism Spectrum Disorders. Vol. 16, International Journal of Molecular Sciences. 2015. 6464–6495 p.
13. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol. 2015;14(11):1109–20.
14. Ruzzo EK, Pérez-Cano L, Jung JY, Wang L kai, Kashef-Haghighi D, Hartl C, et al. Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. Cell. 2019;178(4):850–66.
15. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014 Nov 13;515(7526):209–15.
16. Rylaarsdam L, Guemez-Gamboa A. Genetic Causes and Modifiers of Autism Spectrum Disorder. Front Cell Neurosci. 2019;13(385):1–15.
17. 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;18(10):1795–804.
18. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet. 2014;94(5):677–94.
19. Courchesne E, Pramparo T, Gazestani VH, Lombardo M V., Pierce K, Lewis NE. The ASD Living Biology: from cell proliferation to clinical phenotype. Mol Psychiatry. 2019;24(1):88–107.
20. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485(237–241).
21. 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. 2012;485:246–50.
22. Butte NF, Cai G, Cole SA, Comuzzie AG. Viva la Familia Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population 1–4. Am J Clin Nutr. 2006;84:646–54.
23. Mckenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297–303.
24. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
25. Exome Aggregation Consortium. Analysis of protein-coding genetic variation in 60,706 humans. Nat Publ Gr. 2016;536(6):57.
26. Storey J, Bass A, Dabney A, Robinson D. q-value: Q-value estimation for false discovery rate control. R package version 2.22.0. 2020.
27. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans, Genome Aggregation Database Consortium. Nat . 2020;581:19.
28. Carlson M. KEGG.db: A set of annotation maps for KEGG. R package version 3.2.2.
29. Sanders SJ. Next-generation sequencing in autism spectrum disorder. Cold Spring Harb Perspect Med. 2019;9(8):a026872.
30. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell. 2020;180:1–17.
31. Zhao Y, Kwan KM, Mailloux CM, Lee WK, Grinberg A, Wurst W, et al. LIM-homeodomain proteins Lhx1 and Lhx5, and their cofactor Ldb1, control Purkinje cell differentiation in the developing cerebellum. Proc Natl Acad Sci U S A. 2007;104(32):13182–6.
32. Shawlot W, Behringer RR. Requirement for Lim1 in head-organizer function. Vol. 374,

Nature. 1995. p. 425–30.
33. Kania A, Johnson R, Jessell T. Coordinate roles for LIM Homeobox genes in directing the dorsoventral trajectory of motor axons in vertebrate limb. Cell. 2000;102(2):161–73.
34. Nagamani SCS, Erez A, Shen J, Li C, Roeder E, Cox S, et al. Clinical spectrum associated with recurrent genomic rearrangements in chromosome 17q12. Eur J Hum Genet. 2010;18(3):278–84.
35. Rasmussen M, Vestergaard EM, Graakjaer J, Petkov Y, Bache I, Fagerberg C, et al. 17q12 deletion and duplication syndrome in Denmark—A clinical cohort of 38 patients and review of the literature. Am J Med Genet Part A. 2016;170(11):2934–42.
36. Justice ED, Barnum SJ, Kidd T. The WAGR syndrome gene PRRG4 is a functional homologue of the commissureless axon guidance gene. PLoS Genet. 2017;13(8):e1006865.
37. Yamamoto T, Togawa M, Shimada S, Sangu N, Shimojima K, Okamoto N. Narrowing of the responsible region for severe developmental delay and autistic behaviors in WAGR syndrome down to 1.6Mb including PAX6, WT1, and PRRG4. Am J Med Genet Part A. 2014;164(3):634–8.
38. Krishnan N, Krishnan K, Connors CR, Choy MS, Page R, Peti W, et al. PTP1B inhibition suggests a therapeutic strategy for Rett syndrome. J Clin Invest. 2015;125(8):3163–77.
39. Zhang L, Qin Z, Ricke KM, Cruz SA, Stewart AFR, Chen HH. Hyperactivated PTP1B phosphatase in parvalbumin neurons alters anterior cingulate inhibitory circuits and induces autism-like behaviors. Nat Commun. 2020;11(1):1017.
40. Theodosiou A, Ashworth A. MAP Kinase Phosphatases. Genome Biol. 2002;3(7):reviews3009.1-3009.10.
41. Lee KY, Ahn YM, Joo EJ, Chang JS, Kim YS. The association of DUSP6 gene with schizophrenia and bipolar disorder: Its possible role in the development of bipolar disorder. Mol Psychiatry. 2006;11(5):425–6.
42. Kim SH, Shin SY, Lee KY, Joo EJ, Song JY, Ahn YM, et al. The genetic association of DUSP6 with bipolar disorder and its effect on ERK activity. Prog Neuro-Psychopharmacology Biol Psychiatry. 2012;37(1):41–9.
43. Phelan KD, Shwe UT, Abramowitz J, Birnbaumer L, Zheng F. Critical role of canonical transient receptor potential channel 7 in initiation of seizures. Proc Natl Acad Sci U S A. 2014;111(31):11533–8.
44. Yoon IS, Li PP, Siu KP, Kennedy JL, Macciardi F, Cooke RG, et al. Altered TRPC7 gene expression in bipolar-I disorder. Biol Psychiatry. 2001;50(8):620–6.
45. Lai MC, Kassee C, Besney R, Bonato S, Hull L, Mandy W, et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. The Lancet Psychiatry. 2019;6(10):819–29.
46. Lim CS, Alkon DL. Inhibition of coactivator-associated arginine methyltransferase 1 modulates dendritic arborization and spine maturation of cultured hippocampal neurons. J Biol Chem. 2017;292(15):6402–13.
47. Zhang Y, Hoxha E, Zhao T, Zhou X, Alvarez-Bolado G. Foxb1 Regulates negatively the proliferation of oligodendrocyte progenitors. Front Neuroanat. 2017;11(53):1–12.
48. Luo F, Zhang J, Burke K, Miller RH, Yang Y. The activators of cyclin-dependent kinase 5 p35 and p39 are essential for oligodendrocyte maturation, process formation, and myelination. J Neurosci. 2016;36(10):3024–37.
49. Ouyang L, Chen Y, Wang Y, Chen Y, Fu AKY, Fu WY, et al. p39-associated Cdk5 activity regulates dendritic morphogenesis. Sci Rep. 2020;10(1):1–16.
50. Comi AM, Zimmerman AW, Frye VH, Law PA, Peeden JN. Familial clustering of autoimmune disorders and evaluation of medical risk factors in autism. J Child Neurol. 1999;14(6):388–94.
51. Atladóttir HÓ, Pedersen MG, Thorsen P, Mortensen PB, Deleuran B, Eaton WW, et al. Association of family history of autoimmune diseases and autism spectrum disorders. Pediatrics. 2009;124(2):687–94.

52. Wu S, Ding Y, Wu F, Li R, Xie G, Hou J, et al. Family history of autoimmune diseases is associated with an increased risk of autism in children: A systematic review and meta-analysis. Neurosci Biobehav Rev. 2015;55:322–32.
53. Rom AL, Wu C Sen, Olsen J, Jawaheer D, Hetland ML, Mørch LS. Parental Rheumatoid Arthritis and Autism Spectrum Disorders in Offspring: A Danish Nationwide Cohort Study. J Am Acad Child Adolesc Psychiatry. 2018;57(1):28-32.e1.
54. Zhu Z, Tang S, Deng X, Wang Y. Maternal Systemic Lupus Erythematosus, Rheumatoid Arthritis, and Risk for Autism Spectrum Disorders in Offspring: A Meta-analysis. J Autism Dev Disord. 2020;50(8):2852–9.
55. Braunschweig D, Krakowiak P, Duncanson P, Boyce R, Hansen RL, Ashwood P, et al. Autism-specific maternal autoantibodies recognize critical proteins in developing brain. Transl Psychiatry. 2013;3(7):e277-9.
56. Tsai PH, Yu KH, Chou IJ, Luo SF, Tseng WY, Huang LH, et al. Risk of autism spectrum disorder in children born to mothers with systemic lupus erythematosus and rheumatoid arthritis in Taiwan. Jt Bone Spine. 2018;85(5):599–603.
57. Chen S wei, Zhong X shan, Jiang L na, Zheng X yan, Xiong Y quan, Ma S juan, et al. Maternal autoimmune diseases and the risk of autism spectrum disorders in offspring: A systematic review and meta-analysis. Behav Brain Res. 2016;296:61–9.
58. Kronzer VL, Crowson CS, Sparks JA, Myasoedova E, Davis J. Family history of rheumatologic, autoimmune, and non-autoimmune diseases and risk of rheumatoid arthritis. Arthritis Care Res (Hoboken). 2019;73(2):180–7.
59. Warren RP, Odell JD, Warren WL, Burger RA, Maciulis A, Daniels WW, et al. Strong association of the third hypervariable region of HLA-DR β1 with autism. J Neuroimmunol. 1996;67(2):97–102.
60. Torres AR, Maciulis A, Stubbs EG, Cutler A, Odell D. The transmission disequilibrium test suggests that HLA-DR4 and DR13 are linked to autism spectrum disorder. Hum Immunol. 2002;63(4):311–6.
61. 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. 2006;67(4–5):346–51.
62. Roudier J. Association of MHC and rheumatoid arthritis association of RA with HLA-DR4: The role of repertoire selection. Arthritis Res. 2000;2(3):217–20.
63. Fries JF, Wolfe F, Apple R, Erlich H, Bugawan T, Holmes T, et al. HLA-DRB1 genotype associations in 793 white patients from a rheumatoid arthritis inception cohort: Frequency, severity, and treatment bias. Arthritis Rheum. 2002;46(9):2320–9.
64. Crespi BJ, Thiselton DL. Comparative immunogenetics of autism and schizophrenia. Genes, Brain Behav. 2011;10(7):689–701.

Supplementary Table 1. Genes recurrently accumulating rare genetic variants in ASD population

Genes	phs000482 ASD	phs000298 ASD	phs000298 Ctrl	phs000616 Ctrl	10.1038-ASD	10.1038 Ctrl	All ASD	All Ctrl	ASD cohort	p-value	q-value	Nonsyn SNVs_ASD	Nonsense ASD	Splicing ASD	z-score	pLi
ACTR6	1	7	0	1	3	0	11	1	1;7;3	0.0445	0.3650	10	0	1	1.8929	5.46E-09
ARH1	0	8	0	0	2	0	10	0	0;8;2	0.0205	0.2224	3	0	7	4.5919	1
ATP6V1E2	1	7	0	0	3	0	11	0	1;7;3	0.0138	0.1707	10	1	0	-0.5496	0.0029
B4GALT1	4	7	0	0	2	1	13	1	4;7;2	0.0210	0.2257	10	1	2	0.9757	0.0374
B4GAT1	2	6	0	0	3	1	11	1	2;6;3	0.0445	0.3650	9	2	0	1.9000	0.0100
BSPRY	3	5	0	0	3	1	11	1	3;5;3	0.0445	0.3650	11	0	0	0.1292	1.33E-06
C12orf60	6	9	1	0	1	0	16	1	6;9;1	0.0068	0.1041	16	0	0	-0.0111	0.0483
C17orf102	3	7	0	0	0	0	10	0	3;7;0	0.0205	0.2224	9	1	0	-0.3641	0.0081
C8orf76,ZHX1-C8orf76	2	6	0	0	3	0	11	0	2;6;3	0.0138	0.1707	9	0	2	0	0
CACFD1	2	6	0	0	3	1	11	1	2;6;3	0.0445	0.3650	10	1	0	0.1480	8.06E-04
CARM1	2	7	0	0	2	0	11	0	2;7;2	0.0138	0.1707	2	0	9	4.2514	1
CBR3	2	12	1	0	2	0	16	1	2;12;2	0.0068	0.1041	15	0	1	-0.0345	3.02E-06
CCDC169-SOHLH2,SOHLH2	2	11	1	0	1	0	14	1	2;11;1	0.0144	0.1759	12	0	2	1.0300	0
CCDC175	80	0	0	0	2	1	82	1	80;0;2	1.17E-13	9.45E-12	2	1	79	1.1346	4.92E-09
CCDC85A	3	6	0	0	2	1	11	1	3;6;2	0.0445	0.3650	11	0	0	-0.2144	1.54E-05
CDK5R2	0	12	0	0	2	1	14	1	0;12;2	0.0144	0.1759	14	0	0	2.1699	0.6707
CYB561A3	1	9	0	1	1	0	11	1	1;9;1	0.0445	0.3650	9	0	2	0.8982	0.0110
CYTH3	3	4	0	0	3	0	10	0	3;4;3	0.0205	0.2224	10	0	0	2.5696	0.7225
DUSP6	2	5	0	0	6	0	13	0	2;5;6	0.0064	0.0994	11	0	2	1.2166	0.9760
E2F3	2	13	0	1	2	0	17	1	2;13;2	0.0047	0.0790	16	0	1	1.5902	0.9458
FOXB1	0	6	0	0	5	0	11	0	0;6;5	0.0138	0.1707	11	0	0	1.4834	0.9063
GALNT4,POC1B-GALNT4	3	9	0	1	1	0	13	1	3;9;1	0.0210	0.2257	13	0	0	0.0100	0.2700
HNRNPA1L2	1	10	1	0	0	0	11	1	1;10;0	0.0445	0.3650	11	0	0	-0.0677	0.0188
KLHL24	1	10	1	0	1	0	12	1	1;10;1	0.0305	0.2914	8	0	4	2.8837	0.0017

(Continued)

Genes	phs000482 ASD	phs000298 ASD	phs000298 Ctrl	phs000616 Ctrl	10.1038 ASD	10.1038 Ctrl	All ASD	All Ctrl	ASD cohort	p-value	q-value	Nonsyn SNVs_ASD	Nonsense ASD	Splicing ASD	z-score	pLi
LHX1	3	11	0	0	3	0	17	0	3;11;3	0.0014	0.0292	17	0	0	1.9230	0.0292
LIPF	0	9	0	0	3	1	12	1	0;9;3	0.0305	0.2914	12	0	0	0.4214	7.21E-04
LRRC25	0	79	1	0	1	0	80	1	0;79;1	2.53E-13	1.98E-11	79	1	0	0.8066	0.9450
MEGF9	2	7	0	1	2	0	11	1	2;7;2	0.0445	0.3650	9	0	2	1.4315	0.9881
NSUN3	3	7	1	0	1	0	11	1	3;7;1	0.0445	0.3650	11	0	0	0.3061	9.04E-09
P2RX3	2	14	0	1	2	0	18	1	2;14;2	0.0032	0.0591	18	0	0	1.2711	1.03E-06
PDP1	2	24	0	0	2	1	28	1	2;24;2	7.89E-05	0.0024	28	0	0	1.6189	0.4511
PRRG4	2	32	0	1	1	0	35	1	2;32;1	5.87E-06	2.12E-04	4	0	31	1.3126	0.1128
PTPN1	2	11	1	0	3	0	16	1	2;11;3	0.0068	0.1041	15	1	0	2.7911	0.8129
RAB40B	0	9	0	0	1	0	10	0	0;9;1	0.0205	0.2224	8	1	1	0.4615	8.66E-06
REG3A	0	10	0	1	1	0	11	1	0;10;1	0.0445	0.3650	10	0	1	-0.2025	6.29E-04
REM1	2	8	0	0	2	0	12	0	2;8;2	0.0094	0.1302	11	1	0	0.6719	5.34E-08
RNPC3	7	3	0	0	0	0	10	0	7;3;0	0.0205	0.2224	10	0	0	0.9898	1.96E-04
RUNX2	0	10	0	0	1	0	11	0	0;10;1	0.0138	0.1707	11	0	0	1.6033	0.9488
SOCS4	1	9	0	0	1	0	11	0	1;9;1	0.0138	0.1707	11	0	0	0.6374	1.42E-05
SYNDIG1L	2	9	0	0	0	1	11	1	2;9;0	0.0445	0.3650	11	0	0	0.2128	0.0088
THTPA	1	9	0	1	3	0	13	1	1;9;3	0.0210	0.2257	13	0	0	-0.3336	2.77E-06
TRPC7	0	15	0	0	1	1	16	1	0;15;1	0.0068	0.1041	16	0	0	2.3053	1.95E-04
YIF1B	4	2	0	1	5	0	11	1	4;2;5	0.0445	0.3650	9	0	2	1.4036	3.43E-05
ZNF35	1	9	1	0	2	0	12	1	1;9;2	0.0305	0.2914	10	1	1	1.3593	0.0420

DISCUSSION

It is well accepted that ASD genetic architecture is highly complex and with a high degree of genetic variability shaped by different types of variants and explained by monogenic, polygenic and multifactorial models (49,72). The advent of genomic technologies, such as NGS, has enabled the discovery of thousands (>1000) of genes linked to ASD (71). Most studies have focused on the identification of monogenic causes of ASD and/or highly penetrant variants and have exhaustively examined single-copy regions and coding regions (78,103,226). However, only in around 25-35% of ASD patients the genetic etiology underlying this condition is known (49). This is why, complex regions and non-coding regions, which have been poorly inspected, may contain part of the lost heritability and its exploration can help shed light into the ASD genetic architecture.

In this work, we provide a thorough assessment of complex genomic regions that can contribute to find part of the missing heritability of ASD. On the one hand, 7q11.23 *locus*, which is a well-known region that has been highly associated with ASD. On the other hand, unexplored complex *loci* 8p23.1 and 17q21.31 which bear common ancestral inversions. In addition, we performed genomic and transcriptomic analysis to unravel common pathophysiological mechanisms in ASD.

Complex genomic regions in ASD

The presence of segmental duplications (SDs) predisposes the emergence of chromosomal structural variation such as deletions, duplications and inversions (128). The three complex genomic regions assessed in this thesis (7q11.23, 8p23.1 and 17q21.31) are flanked by blocks of SDs that mediate NAHR events leading to the studied recurrent rearrangements.

Exploration of the 7q11.23 *locus*

The 7q11.23 region is a well-studied *locus*, in which reciprocal copy number variants lead to a paradigmatic pair of neurodevelopmental disorders: Williams-Beuren syndrome and 7q11.23 microduplication syndrome (171). Both

disorders have been associated with ASD. Concretely, 12% of WBS patients present comorbid ASD, representing a 6-10-fold increase in ASD prevalence among WBS subjects compared to controls (196). As for 7q11.23 microduplication syndrome (Dup7), a higher prevalence of ASD (19-40%) has also been reported considering this duplication a susceptibility allele for ASD (84,194). Given the implication of the 7q11.23 *locus* in ASD, we have explored the autistic phenotype through exome sequencing in patients presenting WBS and comorbid ASD as well as transcriptomic analysis in WBS and Dup7 patient-specific iPSCs and derivatives.

We carried out an exome-wide analysis of eight WBS subjects with co-occurring ASD in order to search for second-hit genetic modifiers of the ASD phenotype. Atypical deletions, alteration of flanking genes and common shared haplotypes were discarded as possible *cis*- or *trans*-acting factors in our cohort. Interestingly, we found a parental bias regarding the deletion origin as in 7/8 patients occurred in the paternal allele. This was not observed in a larger cohort of WBS without ASD (n=374), neither in a previous study of WBS subjects with comorbid ASD, where the deletion was originated in the maternal allele in 4/5 patients (195). These findings suggest a possible imprinting effect, but further evidence is needed to confirm it. Also, we detected an over-representation of a synonymous variant (rs12539160) in *MLXIPL*, which was previously associated with ASD in a GWAS study (227). In our cohort, this variant did not reach statistical significance (p=0.076), probably due to the small sample size, and it was found in two individuals (MAF=0.25), representing a 4-8-fold increase in frequency compared to Spanish (0.03) or European population (0.06) (228). We analyzed in detail the frequency of this SNP in non-autistic WBS observing that it was similar to the general population (229), suggesting a possible contribution to ASD risk in WBS with co-morbid ASD.

When assessing second-hit exome-wide variants (SNVs and CNVs), apart from the *de novo* 7q11.23 deletion, six of the patients carried one to three rare genetic variants in candidate genes, with higher burden in females compared with males. The fact that rare deleterious variants in candidate genes were

significantly higher in females than males is consistent with the female protective effect, in which females require a higher genetic load to reach the threshold of developing ASD (101). We detected two small CNVs affecting ASD candidate genes in two patients, including a partial duplication of the last two exons of *SIK1* and a complete heterozygous deletion of *DUSP22*. *De novo* variants in *SIK1* were previously reported in patients with developmental epilepsies and autism features (230). Nevertheless, *DUSP22* has been removed from the SFARI database, no longer supporting a possible contribution of this deletion to the ASD phenotype. We also found LoF and deleterious SNVs in ASD-related genes and functionally constrained genes, highlighting a *de novo* nonsense variant in *UBR5*. Therefore, our results suggest that *trans*-acting factors in the 7q11.23 remaining allele and inherited or *de novo* rare genetic variants elsewhere in the genome may play a role in the susceptibility to ASD in WBS patients.

Next, and to avoid the limitations of the study of neurodevelopment disorders in humans, we used an *in vitro* approach to characterize the 7q11.23 rearrangements. We generated patient-specific iPSCs from fibroblasts from four Dup7 patients (three copies of 7q11.23), four WBS patients (one copy) and two controls (two copies), which were differentiated to NPCs and neurons (iNeus). We presented one of the largest 7q11.23 patient-derived iPSC lines and derivatives, including both duplication and deletion models, reported up to date. The majority of previous iPSC-based transcriptomic studies only included WBS subjects (231–234), while Dup7 patients were studied in very few investigations (235,236). This is why, collaborative efforts in the scientific community, such as conducting a meta-analysis of all 7q11.23 patient-derived iPSCs lines available, would be a powerful tool to overcome sample size limitations.

We appraised genome-wide differential expression using microarray expression analysis in fibroblasts, iPSCs, NPCs and iNeus. Pairwise comparison of the three genotypes in all cell types revealed a total of 494 unique significant DEGs (q-value<0.05). The expression of genes located in the 7q11.23 single-copy region

showed a symmetrically opposite expression in half of them (15/30), in which genes were upregulated in Dup7 and downregulated in WBS lines, and the other half were deregulated in the same direction. However, only four genes (*BAZ1B, BCL7B, CLIP2* and *GTF2IRD1*) were significantly altered (q-value<0.05) when comparing Dup7 or WBS lines against controls. Our results were in line with previous findings from transcriptional studies of 7q11.23 iPSC-based cells in approximately 92-96% of cases depending on the cell type compared (231–233,235). Concretely, *BAZ1B, BCL7B, TBL2, WBSCR22, EIF4H, RFC2, CLIP2* and *GTF2IRD1* have been consistently observed to be deregulated in all studies and cell types. This data suggests that these genes may be more dose sensitive and/or probably fewer *trans* feedback mechanisms are acting to normalize their expression. In addition, we found *MIR590* severely affected by 7q11.23 CNVs as it presented the most extreme log_2FC values both in Dup7 and WBS models, and interestingly, its targets were enriched in ASD risk genes as well as genes linked to WBS (empirical p-value<0.0001).

Enrichment analysis among DEGs and WGCNA analysis revealed pathways and gene-ontology categories related to the hallmark phenotypes of both disorders, to the regulation of gene expression and some categories relevant for the ASD phenotype of Dup7 and/or WBS patients. In fibroblasts, GO categories were involved in the development of the cardiovascular system, skeletal system, renal system and nervous system, which are impaired in both disorders. Similarly, Adamo *et al.* reported resembling results in iPSC lines from Dup7 and WBS individuals (235). ASD-related transcriptional consequences will be discussed in further sections.

The following figure 1 summarizes the main findings in the exploration of 7q11.23 region.

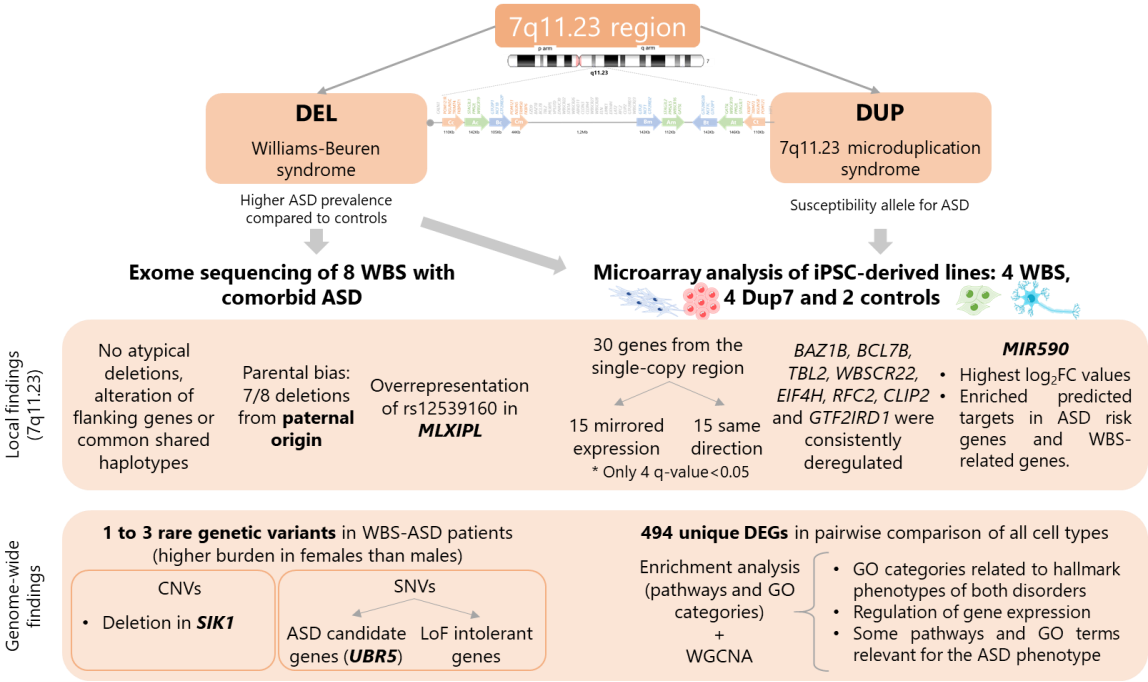


Figure 1. Summary of the main findings in the 7q11.23 region. DUP: duplication, DEL: deletion, DEGs: differentially expressed genes, GO: gene-ontology, WGCNA: weighted gene co-expression analysis, CNVs: copy number variants, SNVs: single nucleotide variants, LoF: loss of function.

Unexplored 8p23.1 and 17q21.31 inversions

The two other complex *loci* addressed in this thesis are 8p23.1 and 17q21.31, where ancestral inversions have been described (139). As inversions are usually copy-neutral changes and in some cases they are difficult to detect and/or genotype, they are not the focus of many studies. However, inversion breakpoints can disrupt genes and alter gene expression as well as predispose to genetic imbalances (142). In fact, some previous evidences related inversions with ASD, by the direct disruption of a gene (*FOXP1)* (237) or as susceptibility factors (136). This is why, we have hypothesized that common inversions could account for part of the missing heritability in ASD.

Up to date, the association between inv8p23.1 and inv17q21.31 with ASD have not been fully covered. Interestingly, in a recent transcriptome-wide association study assessing differential gene expression in ASD, both *loci* showed

transcriptome-wide significant associations (238). Additionally, newer methods have been developed, such as *scoreInvHap* (147), that allowed us to reliably genotype inversions from SNP array data.

We performed a transmission disequilibrium test (TDT) using SNP array data from 3,770 ASD trios and 2,217 control trios of European ancestry. We found a significant over-transmission of both inverted alleles at inv8p23.1 (OR=1.12, p=0.0006) and inv17q21.31 (OR=1.12, p=0.0047) in ASD patients, suggesting that these alleles may act as susceptibility factors for ASD in European population. Besides, the I allele at inv8p23.1 seemed to be associated with verbal ASD subjects and not strict autism, whereas H2-I allele at inv17q21.31 with high IQ and not strict autism as well as it was more prevalent in multiplex families. Nevertheless, the meta-analysis of control samples for the inv17q21.31 showed a tendency for the inverted allele (OR=1.07, p=0.2233) suggesting that 17q21.31 inversion is undergoing adaptive selection in the general population, and it is in line with previous suspicions. Inversion carrier females were showed to have more children and higher genome-wide recombination rates compared to non-carriers in a study with Icelandic population (148). Although inv17q21.31 showed some over-transmission tendency in general population, inverted alleles at 8p23.1 and 17q21.31 may act as susceptibility risk factors for ASD in Europeans, explaining part of the missing heritability in ASD.

To explore the effect of these novel susceptibility ASD factors we conducted an RNA sequencing analysis of peripheral blood in 13 idiopathic ASD patients and 11 controls grouped by inversion genotypes. In the inverted vs non-inverted approach, we discovered differential gene expression for both single-copy genes within the inverted intervals and multiple-copy genes at the SDs. Of particular interest is *KANSL1-AS1*, a long non-coding RNA that was found upregulated in I-allele carriers. Our findings are in agreement with previously reported expression studies of inv17q21.31 (152), whereas in the inv8p23.1 our results are in an opposite direction (144).

Moreover, we carried out a second approach in which we compared ASD subjects carrying the susceptibility allele (inverted status) with all others. Our

results indicated that ASD inverted individuals presented unique ASD transcriptional consequences. Interestingly, we discovered a large number of lncRNAs among the detected DEGs. In particular, 8p23.1 inverted ASD individuals showed 36 DEGs and some of the GO categories discovered in the enrichment analysis were related to actin filament binding and cortical cytoskeleton. For inv17q21.31, we found 19 DEGs in ASD patients carrying the H2-I allele that were enriched for pathways and GO terms related to immune system and inflammatory processes, mainly involving upregulated genes. These results will be further discussed in following sections.

In light of the altered expression of local genes, we thought that the inversion could directly affect the expression of such genes by means of the disruption of topologically associated domains (TADs). Curiously, both 8p23.1 and 17q21.31 regions contain TADs (239). There is evidence that structural genomic rearrangements can modify the 3D genomic organization by disrupting TADs and cause regulatory gain and/or loss of function affecting gene expression (240,241). For instance, in developmental disorders, ~7% of balanced chromosomal anomalies produce TAD disruption (242) and rare inherited structural variants disrupting cis-regulatory elements (CRE-SVs) seem to increase ASD risk (243). In the specific case of inversions, they are only suspected to influence gene regulation when disrupting a coding gene or TAD boundary (241). Since in both inversion intervals TADs are present, they could be disrupted by inversion breakpoints suggesting that the gene expression alteration can be explained by a direct inversion-mediated mechanism. In order to experimentally validate this suspicion, chromosome conformation capture (3C-based) techniques, such as Hi-C, should be performed in future studies.

Despite the detection of altered local gene expression, the majority of DEGs in our analysis were located elsewhere in the genome. We have speculated that a possible mechanism to explain the expression changes of these genes could be a disruption of non-homologous chromosomal contacts (NHCCs), where gene regulatory regions from one chromosome can affect genes located in other chromosomes (244), There is a striking lack of information concerning the

NHCCs mechanism, but it is known that lncRNAs are involved in its 3D organization (240) and higher interaction frequency between chromosomes 14-17 and 19-22 has been described (245). This evidence lends support to our hypothesis as we showed not only a remarkable proportion of DEGs in chromosomes 14 and 17, but also a high number of deregulated lncRNAs. Similarly to what has been described for TADs, chromosomal variants can disrupt and reorganize the NHCCs network, such as alteration of chromatin contacts by CNVs at 16p11.2 *locus* that have been associated with an ASD phenotype (240,246).

An overview of the principal findings from the two unexplored complex regions (8p23.1 and 17q21.31) can be found in figure 2.

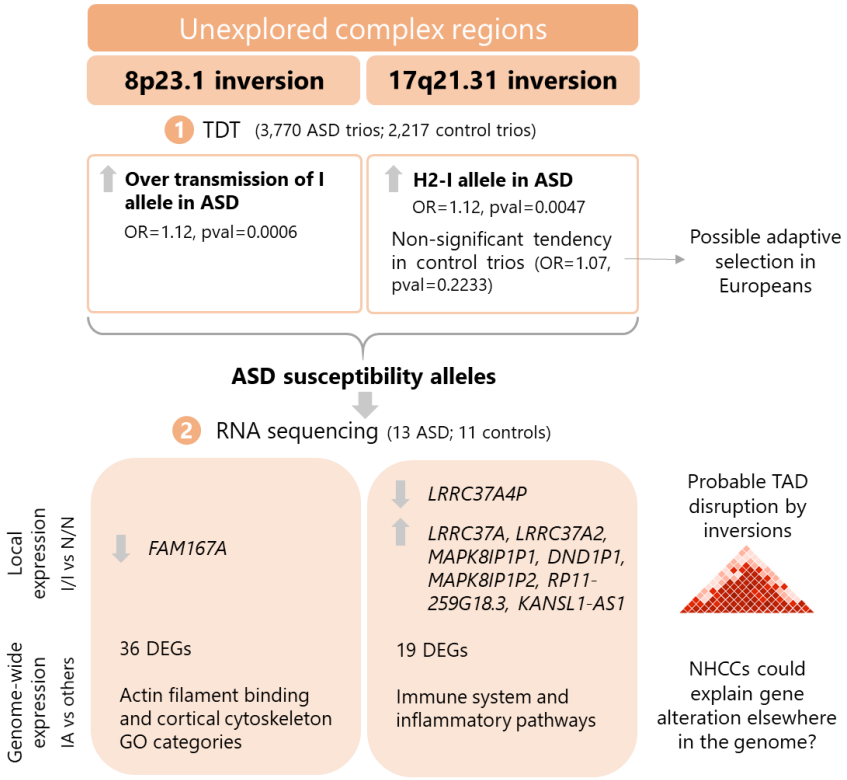


Figure 2. Outline of the principal results in 8p23.1 and 17q21.31 unexplored regions. TDT: transmission disequilibrium test, OR: odds ratio, IA: inverted ASD patients, DEGs: differentially expressed genes; GO: gene-ontology, TAD: topologically associated domains; NHCCs: non-homologous chromosomal contacts.

Pathophysiological mechanisms in ASD

In this work, we have applied different strategies, including transcriptome sequencing, microarray expression and exome sequencing, to deepen in the identification of deregulated pathophysiological mechanisms in ASD. We have included patients with ASD susceptibility factors (7q11.23 rearrangements, inv8p23.1 and inv17q21.31) and idiopathic patients considering a multiple-hit model.

Deregulated functional pathways

A common procedure performed in this work is the use of enrichment pathway and gene-ontology (GO) analysis in order to identify which pathways are over-represented by altered genes.

The ASD-related transcriptomic consequences derived from *in vitro* cellular models from Dup7 individuals were obtained through enrichment analysis as well as weighted gene co-expression analysis (WGCNA). First, DEGs from Dup7 fibroblasts only revealed one pathway significantly enriched that was Arf6 trafficking events (q-value=0.0313). This pathway is implicated in endosomal membrane trafficking and actin remodelling, which are tightly connected processes with important functions for neuronal development as well as for synaptic transmission and plasticity (247,248). Also, a recent study described a patient with ID, autism and epilepsy carrying a mutation in *IQSEC2* (IQ motif and Sec7 domain ArfGEF 2) that resulted in increased production of ARF6 (249). Therefore, impairments in the Arf6 trafficking events pathway could partially contribute to the autistic phenotype of Dup7 individuals.

Second, WGCNA is a tool for studying biological networks and identifying co-expression modules associated with a trait of interest. We performed WGCNA in iPSCs and NPCs lines from Dup7 and WBS individuals and looked for co-expression modules related to the number of 7q11.23 copies. In iPSCs, this approach uncovered three co-expression modules significantly correlated with 7q11.23 genotype. Of particular interest was $M1_{iPSC}$ module that was directly

related to the number of 7q11.23 copies, and thus mainly contained genes upregulated in Dup7 and downregulated in WBS. It was enriched in genes involved in cholesterol biosynthesis and ion transport. Interestingly, cholesterol homeostasis is crucial for neural development and synaptic function and altered cholesterol levels have been reported in some ASD patients suggesting a possible role in ASD pathogenesis (250–252). Lipid rafts disarrangements, in which cholesterol is a key component, could lead to synaptic dysfunction providing a link between cholesterol metabolism and ASD (250). Given the fact that low-normocholesterolemia has been observed among WBS patients (253), the alteration in cholesterol biosynthesis may be more relevant for the autistic phenotype of Dup7 individuals.

WGCNA analysis in NPCs unveiled five co-expression modules significantly associated with 7q11.23 dosage. Remarkably, $M41_{NPC}$ module was inversely related to the number of 7q11.23 copies and was enriched in genes related to the regulation of glutamate receptor signaling pathway. Multiple evidence supports a glutamate dysfunction in ASD pathophysiology: 1) increased levels of glutamate in blood and brain of ASD patients compared to controls (254); 2) alterations in members of the glutamatergic system by microarray analysis, such as higher mRNA levels of AMPA-type glutamate receptors and glutamate transporters in post-mortem brain tissues of ASD subjects (214,255); 3) various strong ASD risk genes are involved in glutamatergic synapses, including *GRIN2B, GRIA2* or *SLC1A2* (256); and 4) genes located in ASD-specific CNVs were enriched in glutamatergic synapse in multiple studies (88,257). Therefore, the glutamate receptor signaling could contribute to the etiology of the autistic phenotype of both Dup7 and WBS patients.

Lastly, enriched pathways and GO categories in DEGs from Dup7 iNeus were related to neuronal processes. Some of the top-ranking signaling pathways included transmission across chemical synapses, neuronal system, neurotransmitter release cycle and GABAergic, serotonergic, glutamatergic and dopaminergic synapse, among others. These results further support the role of the glutamatergic system in the pathogenesis of ASD. Taken together, these

findings were likely to indicate an alteration of synaptic function in Dup7 neurons, a process consistently impaired in ASD (209).

In the RNA sequencing from ASD patients and controls grouped by inv8p23.1/inv17q21.31 genotypes, we used a second approach in which we looked for inverted ASD individuals' specific transcriptional consequences. For 8p23.1, we discovered GO categories mainly related to actin filament binding and cortical cytoskeleton. It is well accepted that the regulation of cytoskeletal dynamics is essential for dendritic spine morphology and formation (258). Alterations in actin filaments have been observed in both stem cells derived from ASD patients and in an autism-like mouse model (259,260). In addition, several ASD risk genes identified by genetic and gene expression studies are involved in cytoskeleton regulation (261). In fact, we have found Arf6 trafficking events pathway significantly enriched among DEGs in Dup7 fibroblasts, further supporting a possible contribution of actin cytoskeleton regulation in ASD.

Regarding inv17q21.31, immune-inflammatory pathways were enriched in ASD inverted patients, mainly including upregulated genes. This finding points out the implication of a deregulation of immune-inflammatory pathways in ASD pathogenesis, which reinforces previous reported results. Upregulation of genes linked to the immune and inflammatory response in ASD has been highlighted in several microarray and RNA sequencing investigations from peripheral tissues and brain (216,218,219,262–264). Some studies reported both upregulation and downregulation of genes involved in immune-inflammatory pathways (265–267) and controversially, two analysis found decreased peripheral immune-inflammatory functions in ASD (221,268). In spite of the contradictory findings, evidence clearly points towards an immune imbalance in ASD patients. Alteration of various immune cell type levels in blood and brain as well as higher frequency of infections or allergies have been reported in ASD patients. Other evidence relies on the fact that maternal infections and maternal immune system vulnerabilities during pregnancy are considered ASD risk factors (269–271). Given that immune system plays important roles in neurodevelopment and in normal brain functions, immune dysfunction could

alter neuronal synaptic function (272). In addition, a strong neuroinflammation has been observed in ASD patients, which could lead to synaptic damage (269,273). All in all, our findings support the implication of immune-inflammatory pathways in ASD.

Among maternal immune system vulnerabilities, family history of autoimmune diseases has been consistently observed to increase ASD risk in offspring (53). We have performed a meta-analysis of exomic data in 1,078 idiopathic ASD patients and 747 control individuals with the purpose of identifying common deregulated pathways among ASD population due to the accumulation of genes that recurrently harbor rare genetic variants. Only the Rheumatoid arthritis (RA) pathway was found significantly enriched (q-value=0.0138) in ASD patients under a multiple-hit model. Curiously, this pathway was also deregulated among DEGs from ASD subjects carrying the 17q21.31 inverted alleles. Data on autoimmune diseases in patients with ASD is limited, but no comorbidity of RA in ASD individuals has been described so far. Several studies reported an increased number of ASD subjects with relatives suffering from juvenile or adult RA compared to controls and that maternal family history of RA was associated with a higher risk of ASD in offspring (274–278). It is thought that maternal autoantibodies can act against fetal brain tissue affecting neurodevelopment (279,280). However, there are some controversial findings not supporting a significant association between maternal RA and ASD risk in offspring (281,282). On the other hand, this association has also been found in the opposite direction as family history of non-autoimmune diseases, including autism, was related to increased RA risk (283). Furthermore, allelic products of certain genes of the MHC have been identified with a significantly higher frequency in both ASD and RA patients (284–288), suggesting common genetic pathways between ASD and RA, which are possibly explained by a model of pleiotropic effects from risk alleles (289). Our data furthers supports a common genetic background between ASD and RA.

Taken all these results together, we can observe that relevant pathways and GO categories identified in our study by different approaches converge towards a

possible synaptic dysfunction as it is shown in figure 3. It is tempting to speculate that the alteration of this biological process could be a possible common pathophysiological mechanism implicated in ASD contributing to the ASD phenotype.

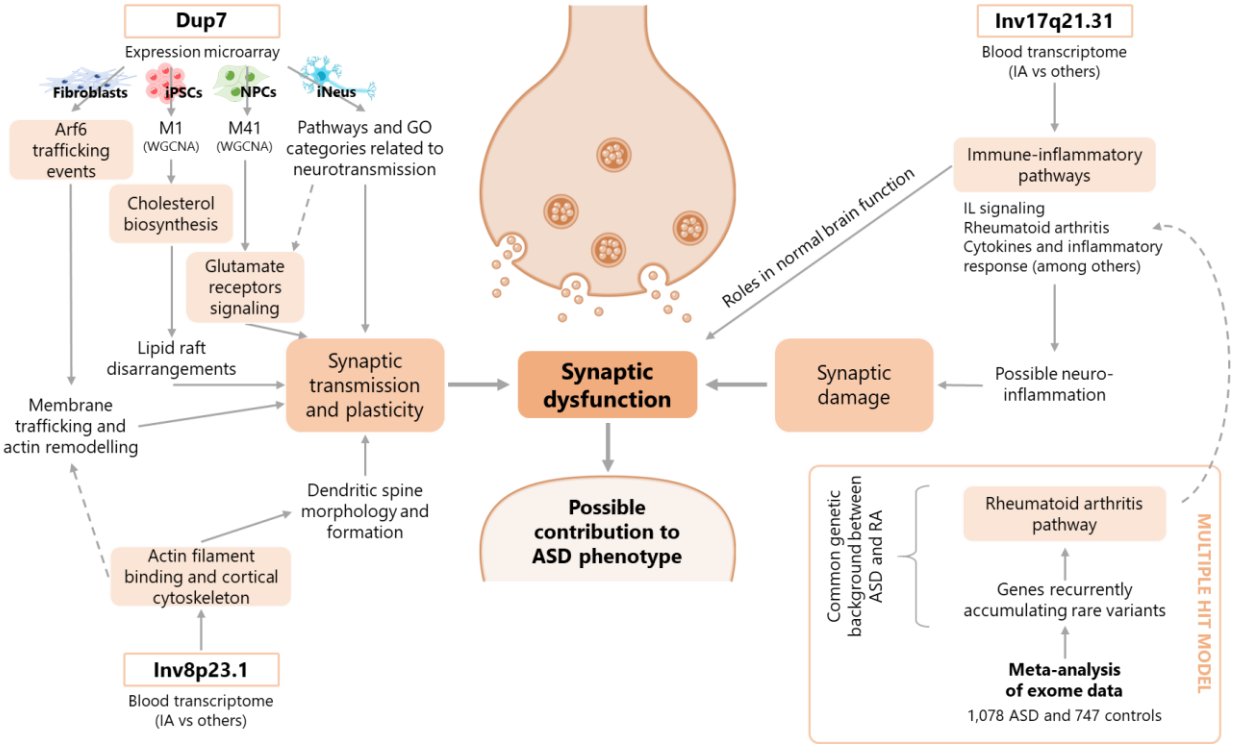


Figure 3. Common deregulated functional pathways in ASD patients points towards a probable synaptic dysfunction.

Possible new ASD candidates

In this work, we have collectively identified several strong ASD risk genes previously reported in SFARI database (*UBR5, PTGS2* and *MYH10*) as well as some genes related to ASD to a lesser extent in prior studies (*IL1B, LHX1, PRRG4* and *PTPN1*). In addition, we discovered further genes that could be proposed as novel ASD susceptibility genes given its implication in ASD-comorbid conditions or its biological relevance in central nervous system.

Among genes previously linked to ASD, we detected a *de novo* nonsense mutation in *UBR5* (ubiquitin-protein ligase E3 component N-recognin 5) in a

severely affected female with WBS and comorbid ASD. *De novo* variants, three missense and one LoF, have been previously described in ASD subjects (106,109,119,290). Regarding *PTGS2* (prostaglandin-endoperoxide synthase 2) and *MYH10* (myosin heavy chain 10), both genes were found significantly upregulated in ASD patients carrying 17q21.31 and 8p23.1 inverted alleles, respectively. Common variants in *PTGS2* have been associated with ASD (291), whereas *de novo* rare variants and mosaic variants in *MYH10* have been observed in ASD patients (73,118,292).

Furthermore, our results revealed four additional genes tied with ASD in a minor degree, as they are not reported in SFARI. On the one hand, we identified a significant overexpression of *IL1B* (interleukin 1 beta) among ASD individuals' carriers of the H2-I allele of inv17q21.31. In agreement with these results, increased serum levels of IL-1β were observed in ASD patients compared to controls, along with other cytokines (293,294) and polymorphisms in this gene significantly elevated ASD risk (294,295). On the other hand, *LHX1* (LIM homeobox 1), *PRRG4* (proline rich and gla domain 4) and *PTPN1* (protein tyrosine phosphatase non-receptor type 1) were genes recurrently mutated in ASD population in the meta-analysis of exome data. *LHX1* has been proposed as a candidate gene for the neurocognitive phenotype of 17q12 duplications and deletions, as both are characterized by autistic features (296,297). Similarly, *PRRG4* has been suggested to be responsible for the ASD phenotype of the WAGR syndrome (OMIM #194072) (298). Lastly, *PTPN1* encodes PTP1B protein, the activation of which appeared to trigger autism-like behaviors in a murine model (lacking *Lmo4* in interneurons) and its inhibition ameliorate neural and behavioral symptoms in RTT mice model (299,300). All in all, our data further sustains a significant role of these above-mentioned genes in autism providing new evidence to support its implication in the ASD pathogenesis.

Interestingly, our work has also uncovered several genes that could have a potential role in ASD. In the iPSC-based microarray expression study from Dup7 and WBS patients, we highlighted *MIR590* as a novel compelling candidate to regulate the expression of other genes relevant for both disorders. This miRNA

has been associated with cardiomyocyte proliferation, anti-apoptotic functions in atherosclerosis and tumor suppressor activity in different cancers (301–304). Investigating the genes that could be post-transcriptionally regulated by *MIR590*, we found that its predicted targets were enriched in ASD risk genes and genes previously linked to WBS (empirical p-value<0.0001). Nowadays, miRNAs are drawing more attention in the field of ASD as growing evidence from expression profiling studies supports a miRNA dysregulation in ASD patients (305). However, there is a lack of consensus between the differentially expressed miRNAs, and in fact, *MIR590* has not been reported before (305,306). A recent meta-analysis only found 7 miRNAs associated with ASD by more than one approach (genetic association, expression profiling and/or functional research) (307). Some of the recurrently altered miRNAs in ASD were miR-23a, miR-132 and miR-146b, and their predicted targets were also enriched for ASD susceptibility genes (308). Experimental confirmation of predicted *MIR590* targets is needed, and we point out the possible value of this miRNA as a therapeutic target for 7q11.23 CNV patients.

Other non-coding elements gaining attention in ASD are lncRNAs. We identified a substantial proportion of differentially expressed lncRNAs as an ASD specific signature for both 17q21.31 and 8p23.1 candidate inversions. Similarly, several transcriptomic analyses have detected many differentially expressed lncRNAs in post-mortem brain tissues and peripheral blood from ASD patients (219,222,223), including *RPS10-AS1* (309), *LINC00693* (219), *DISC2* and *PRKAR2A-AS1* (310), among others. However, the lncRNAs detected in our analysis did not overlap with previously reported. The implication of lncRNAs in synaptogenesis, neurogenesis and GABAergic interneuron functions in brain development might explain the functional relation between ASD and lncRNAs, as these functions have been found impaired in ASD probands (224). After searching the specific tissue expression and the nearest neighbouring genes looking for possible *cis*-regulatory mechanisms, we propose *KANSL1-AS1* and *LINC00938* as new candidate lncRNAs that could play a role in ASD pathogenesis, as both are brain-expressed and their nearest neighbouring

genes are found in SFARI (*KANSL1* and *ARID2*). Our results further advocate for an involvement of lncRNAs in ASD as well as in regulation of gene expression.

Regarding protein-coding genes, we identified new ASD susceptibility genes that were either differentially expressed in ASD 8p23.1 inverted subjects (*ATP13A1, NR4A3* and *SLC12A6*) or recurrently mutated in idiopathic ASD patients under a multiple-hit model (*DUSP6* and *TRPC7*). We support a potential role of these genes in ASD pathogenesis given its implications in other neuropsychiatric disorders, such as bipolar disorder (BD) or schizophrenia (SCZ), which are more prevalent in ASD patients and often observed as co-occurring conditions (46,47). Also, neuropsychiatric disorders have been demonstrated to share a common genetic background with ASD (311,312).

While *ATP13A1* (ATPase 13A1) was downregulated in ASD probands carrying the I-allele of inv8p23.1 compared with all others, both *NR4A3* (nuclear receptor subfamily 4 group A member 3) and *SLC12A6* (solute carrier family 12 member 6) were overexpressed. *ATP13A1* has been suggested to act in early neuronal mice development as its highest expression was in the peak of neurogenesis, and recently has been proposed as a novel candidate for ID (313,314). The transcription factor *NR4A3* appears to mediate neuronal survival and axon guidance in hippocampal development, among other functions (315). Besides, it was involved in the nicotine addiction observed in SCZ and BD (316). Rare variants in *SLC12A6,* which is an electroneural K^+-Cl^- cotransporter, have been associated with BD and SCZ and recessive inherited mutations are the cause of Anderman syndrome (OMIM #218000) (317–319).

Other particularly compelling candidates are *DUSP6* (dual specificity phosphatase 6) and *TRPC7* (transient receptor potential cation channel subfamily C member 7), which significantly accumulated rare genetic variants in ASD population compared to controls and both of them have been associated with BD. The 12q22-23 region, where *DUSP6* is located, is considered a candidate susceptibility locus for major psychiatric disorders and several SNPs in this gene showed positive association with female patients with BD (320,321). Given its potential role as a calcium channel, *TRPC7* has been involved in the

altered Ca^{2+} homeostasis exhibited in some BD patients as it was found significantly reduced in these probands (322). In addition, a study using *TRPC7* KO mice strongly implicated this gene in the initiation of seizures (323).

The following table summarizes the set of genes relevant for ASD found in this thesis:

	Gene	Our data	Previous evidence associated with ASD
Strong ASD candidates	*UBR5*	*De novo* nonsense mutation in a female with WBS and comorbid ASD (p.Arg633*)	
	PTGS2	Upregulated in 17q21.31 inverted ASD patients ($log_2FC=3.79$, q-value=0.0057)	Reported in SFARI as ASD candidate genes
	MYH10	Upregulated in 8p23.1 inverted ASD patients ($log_2FC=1.57$, q-value=0.016)	
Related to ASD in minor degree	*IL1B*	Upregulated in 17q21.31 inverted ASD patients ($log_2FC=3.49$, q-value=0.0057)	Increased serum levels of IL-1B in ASD patients (293,294); polymorphisms in this gene significantly elevated ASD risk (294,295)
	LHX1	Recurrently mutated in ASD population (17 ASD and 0 controls, q-value=0.029)	Candidate gene for the neurocognitive profile of 17q12 CNVs, including ASD in both cases (296,297)
	PRRG4	Recurrently mutated in ASD subjects (35 ASD and 1 control, q-value=2.12E-04)	Responsible for the ASD phenotype of the WAGR syndrome (298)
	PTPN1	Recurrently mutated in ASD population (16 ASD and 1 controls, p-value=0.007)	Activation of the encoded protein seems to trigger autism-like behaviours in a murine model (299); inhibition of these protein improves neural and behavioral symptoms in RTT mice model (300)
Possible novel ASD candidate genes	*MIR590*	Mirror expression between Dup7 and WBS lines; significant in Dup.vs.Del comparison in iPSCs, NPCs and iNeus	Related to cardiomyocyte proliferation and different types of cancers; miRNA deregulation is a consistent finding in ASD expression profiling studies (307)
	KANSL1-AS1	Upregulated in 17q21.31 inverted ASD patients ($log_2FC=2.94$, q-val=1.56E-05)	High number of differentially expressed lncRNAs in post-mortem brain tissues and peripheral blood from ASD patients (219,222,223)
	LINC00938	Downregulated in 17q21.31 inverted ASD patients ($log_2FC=-0.56$, q-val=0.047)	High number of differentially expressed lncRNAs in post-mortem brain tissues and peripheral blood from ASD patients (219,222,223)
	ATP13A1	Downregulated in 8p23.1 inverted ASD patients ($log_2FC=-0.57$, q-value=0.003)	May act in early neuronal mice development (313); novel candidate for ID (314)
	NR4A3	Upregulated in ASD patients carrying the I-allele of inv8p23.1 ($log_2FC=4.98$, q-value= 0.001)	May mediate neuronal survival and axon guidance in hippocampal development (315); implicated in nicotine addiction in SCZ and BD (316)
	SLC12A6	Upregulated in 8p23.1 inverted ASD patients ($log_2FC=1.14$, q-value=0.049)	Potassium-chloride cotransporter associated with BD and SCZ by rare variants (317,318)
	DUSP6	Recurrently mutated in ASD population (13 ASD and 0 controls, p-value=0.006)	Located in a candidate *loci* for major psychiatric disorders (320); several SNPs positively associated with female BD patients (321)
	TRPC7	Recurrently mutated in ASD population (16 ASD and 1 control, p-value=0.007)	Downregulated in BD patients with a probable role in calcium homeostasis (322); may initiate seizures in a KO mice model (323)

Table 1. Overview of ASD risk genes as well as novel ASD candidate genes detected in this work.

Concluding remarks

The complex genetic architecture of ASD gives rise to the need of exploring complex genomic regions in order to find part of the missing heritability in ASD. Inverted alleles at the common inversions 8p23.1 and 17q21.31 may act as susceptibility risk factors for ASD in European population. By using different strategies, we have unraveled several pathways and GO categories, including cholesterol biosynthesis, actin cytoskeletal dynamics, glutamate receptor signaling and immune-inflammatory pathways, that could partially contribute to the underlying molecular mechanism of ASD patients with several susceptibility variants (Dup7, inv8p23.1 and inv17q21.31) and also in idiopathic patients under a multiple-hit model. We further support the implication of several ASD-related genes such as *UBR5, PTGS2* or *PRRG4*. In addition, our study also points to a potential role of novel ASD candidate risk genes (*NR4A3*, *SLC12A6, ATP13A1*, *DUSP6* and *TRPC7*) due to its contribution to other neuropsychiatric disorders. All in all, the work presented in this thesis will contribute to a better understanding of pathways and possible pathophysiological mechanisms implicated in ASD, which can ultimately be targeted to potentiate the development of therapeutic approaches in these patients.

CONCLUSIONS

Conclusions

1. The assessment of second-hit modifiers in WBS patients with comorbid ASD suggested that *trans*-acting factors in the 7q11.23 remaining allele and inherited or *de novo* rare genetic variants elsewhere in the genome may play a role in the susceptibility to ASD in WBS.

2. Patient-derived iPSCs lines from fibroblasts are a good approach to overcome the limitations to study neurodevelopmental disorders in humans such as Dup7 and WBS.

3. Expression analysis of *in vitro* 7q11.23-CNVs cellular models identified a regular gene expression pattern of 7q11.23 single-copy genes with a highly consistent upregulation in Dup7 lines and downregulation in WBS lines of *BAZ1B, BCL7B, TBL2, WBSCR22, EIF4H, RFC2, CLIP2* and *GTF2IRD1*. Interestingly, *MIR590* could be a new compelling candidate to regulate the expression of genes relevant for both Dup7 and WBS disorders.

4. Integrative transcriptomic analysis of *in vitro* 7q11.23-CNVs cellular models revealed pathways and gene-ontology categories relevant for the ASD phenotype of Dup7 patients including actin remodelling and membrane trafficking, cholesterol metabolism and glutamate signaling among other biological processes included in neurotransmission.

5. TDT in ASD and controls showed an overtransmission of the 8p23.1 and 17q21.31 inverted alleles suggesting that they may act as novel susceptibility factors for ASD in European population. Also, inv17q21.31 might be going through adaptive selection in the general population.

6. Inversions have a direct effect on gene expression patterns both locally (region around inversion) and elsewhere in the genome that could be explained by the disruption of TADs and NHCCs.

7. Transcriptomic results showed a large number of deregulated lncRNAs in ASD inverted patients indicating that gene regulation could have an important role in ASD, and particularly implicating *KANSL1-AS1* and *LINC00938*.

8. ASD carriers of the susceptibility alleles (inv8p23.1/inv17q21.31) presented differential expression of neuronal function-related genes that we propose as novel ASD candidate genes including *ATP13A1, NR4A3* and *SLC12A6*.

9. Enrichment analysis of DEGs from ASD patients carrying the I-alleles highlighted the implication of a deregulation of immune-inflammatory pathways and actin cytoskeletal regulation in ASD pathogenesis.

10. Meta-analysis of exome data from ASD patients and controls uncovered several genes recurrently accumulating rare genetic variants in ASD population, which could contribute to ASD risk in a multiple hit model. We advocate for an involvement of *LHX1, PRRG4* and *PTPN1* in ASD pathogenesis and highlight novel candidate genes such as *DUSP6* and *TRPC7*.

11. Rheumatoid arthritis signaling pathway was significantly enriched in ASD probands compared to controls further supporting a common genetic background between ASD and RA.

REFERENCES

References

1. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, fifth edition. Psychiatric News. Arlington, VA: American Psychiatric Publishing; 2013. 1–1 p.
2. Kanner L. Autistic Disturbances of Affective Contact. Nerv Child. 1943;2(3):217–50.
3. Asperger H. Die "Autistischen Psychopathen" im kindesalter. Arch Psychiatr Nervenkr. 1944;117(1):76–136.
4. Wing L. Asperger's syndrome: A clinical account. Psychol Med. 1981;11(1):115–29.
5. Baron-Cohen S. Leo Kanner, Hans Asperger, and the discovery of autism. Lancet. 2015;386(10001):1329–30.
6. Lotter V. Epidemiology of autistic conditions in young children. Soc Psychiatry. 1966;1(3):124–35.
7. Matson JL, Kozlowski AM. The increasing prevalence of autism spectrum disorders. Res Autism Spectr Disord. 2011;5(1):418–25.
8. Hansen SN, Schendel DE, Parner ET. Explaining the increase in the prevalence of autism spectrum disorders: The proportion attributable to changes in reporting practices. JAMA Pediatr. 2015;169(1):56–62.
9. Baxter AJ, Brugha TS, Erskine HE, Scheurer RW, Vos T, Scott JG. The epidemiology and global burden of autism spectrum disorders. Psychol Med. 2015;45(3):601–13.
10. Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, Marcín C, et al. Global Prevalence of Autism and Other Pervasive Developmental Disorders. Autism Res. 2012;5(3):160–79.
11. Poovathinal SA, Anitha A, Thomas R, Kaniamattam M, Melempatt N, Anilkumar A, et al. Global Prevalence of Autism: A Mini-Review . SF J Autism. 2018;2:1.
12. Morales-Hidalgo P, Roigé-Castellví J, Hernández-Martínez C, Voltas N, Canals J. Prevalence and Characteristics of Autism Spectrum Disorder Among Spanish School-Age Children. J Autism Dev Disord. 2018;48(9):3176–90.
13. Pérez-Crespo L, Prats-Uribe A, Tobias A, Duran-Tauleria E, Coronado R, Hervás A, et al. Temporal and Geographical Variability of Prevalence and Incidence of Autism Spectrum Disorder Diagnoses in Children in Catalonia, Spain. Autism Res. 2019;12(11):1693–705.

14. Lyall K, Croen L, Daniels J, Fallin MD, Ladd-Acosta C, Lee BK, et al. The Changing Epidemiology of Autism Spectrum Disorders. Annu Rev Public Health. 2017;38:81–102.

15. Maenner MJ, Shaw KA, Baio J, Washington A, Patrick M, DiRienzo M, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. MMWR Surveill Summ. 2020;69(4):1–12.

16. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of autism spectrum disorder among children aged 8 Years - Autism and developmental disabilities monitoring network, 11 Sites, United States, 2014. MMWR Surveill Summ. 2018;67(6).

17. Brugha TS, McManus S, Bankart J, Scott F, Purdon S, Smith J, et al. Epidemiology of autism spectrum disorders in adults in the community in England. Arch Gen Psychiatry. 2011;68(5):459–66.

18. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. J Am Acad Child Adolesc Psychiatry. 2017;56(6):466–74.

19. Werling DM, Geschwind DH. Sex differences in autism spectrum disorder. Curr Opin Neurol. 2013;26(2):146–53.

20. Russell G, Steer C, Golding J. Social and demographic factors that influence the diagnosis of autistic spectrum disorders. Soc Psychiatry Psychiatr Epidemiol. 2011;46(12):1283–93.

21. Dworzynski K, Ronald A, Bolton P, Happé F. How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? J Am Acad Child Adolesc Psychiatry. 2012;51(8):788–97.

22. Tierney S, Burns J, Kilbey E. Looking behind the mask: Social coping strategies of girls on the autistic spectrum. Res Autism Spectr Disord. 2016;23:73–83.

23. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. 4th edition. Washington, CA: American Psychiatric Publishing; 2000.

24. Mazurek MO, Lu F, Symecko H, Butter E, Bing NM, Hundley RJ, et al. A Prospective Study of the Concordance of DSM-IV and DSM-5 Diagnostic Criteria for Autism Spectrum Disorder. J Autism Dev Disord. 2017;47(9):2783–94.

25. Wiggins LD, Rice CE, Barger B, Soke GN, Lee L-C, Moody E, et al. DSM-5 criteria for autism spectrum disorder maximizes diagnostic sensitivity and specificity in preschool children. Soc Psychiatry Psychiatr Epidemiol. 2019;54(6):693–701.

26. WHO. International Statistical Classification of Diseases and Related Health Problems - 10th Revision (ICD-10). 2019.

27. WHO. International statistical classification of diseases and related health problems - 11th revision (ICD-11). 2019.

28. Lord C, Risi S, DiLavore PS, Shulman C, Thurm A, Pickles A. Autism from 2 to 9 years of age. Arch Gen Psychiatry. 2006;63(6):694–701.

29. Yates K, Le Couteur A. Diagnosing autism/autism spectrum disorders. Paediatr Child Heal (United Kingdom). 2016;26(12):513–8.

30. Delahunty C. Developmental delays and autism: Screening and surveillance. Cleve Clin J Med. 2015;82(S1):S29–32.

31. Johnson CP, Myers SM, Lipkin PH, Cartwright JD, Desch LW, Duby JC, et al. Identification and evaluation of children with autism spectrum disorders. Pediatrics. 2007;120(5):1183–215.

32. Palmer N, Beam A, Agniel D, Eran A, Manrai A, Spettell C, et al. Association of sex with recurrence of autism spectrum disorder among siblings. JAMA Pediatr. 2017;171(11):1107–12.

33. Filipek PA, Accardo PJ, Ashwal S, Baranek GT, Otr L, Cook EH, et al. Practice parameter: screening and diagnosis of autim. Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. Neurol. 2000;55(4):468–79.

34. Robins DL, Casagrande K, Barton M, Chen C-MA, Dumont-Mathieu T, Fein D. Validation of the Modied Checklist for Autism in Toddlers, Revised With Follow-up (M-CHAT-R / F). Pediatrics. 2014;133(1):37–45.

35. Hodges, H., Fealko C, Soares N. Autism spectrum disorder: Definition, epidemiology, causes, and clinical evaluation. Transl Pediatr. 2019;9(S1):S55–65.

36. Ellerbeck K, Smith C, Courtemanche A. Care of children with autism spectrum disorder. Prim Care - Clin Off Pract. 2015;42(1):85–98.

37. Zwaigenbaum L, Bauman ML, Stone WL, Yirmiya N, Estes A, Hansen RL, et al. Early Identification of Autism Spectrum Disorder: Recommendations for Practice and Research. Pediat. 2019;136:S10–40.

38. 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.

39. Lai MC, Lombardo M V., Baron-Cohen S. Autism. Lancet. 2014;383(9920):896–910.

40. Davidovitch M, Levit-Binnun N, Golan D, Manning-Courtney P. Late Diagnosis of Autism Spectrum Disorder After Initial Negative Assessment

by a Multidisciplinary Team. J Dev Behav Pediatr. 2015;36(4):227–34.

41. Bertelli MO. ASD and Intellectual Disability. In: Psychopathology in Adolescents and Adults with Autism Spectrum Disorders. Springer International Publishing; 2019. p. 111–30.

42. Lukmanji S, Manji SA, Kadhim S, Sauro KM, Wirrell EC, Kwon CS, et al. The co-occurrence of epilepsy and autism: A systematic review. Epilepsy Behav. 2019;98:238–48.

43. Brambilla P, Hardan A, Ucelli Di Nemi S, Perez J, Soares JC, Barale F. Brain anatomy and development in autism: Review of structural MRI studies. Brain Res Bull. 2003;61(6):557–69.

44. Kostiukow A, Poniewierski P, Daroszewski P, Samborski W. Gastrointestinal disorders in children with autism spectrum disorder. Pol Med J. 2020;48(283):69–72.

45. Mannion A, Leader G, Healy O. An investigation of comorbid psychological disorders, sleep problems, gastrointestinal symptoms and epilepsy in children and adolescents with Autism Spectrum Disorder. Res Autism Spectr Disord. 2013;7(1):35–42.

46. Rosen TE, Mazefsky CA, Vasa RA, Lerner MD. Co-occurring psychiatric conditions in autism spectrum disorder. Int Rev Psychiatry. 2018;30(1):40–61.

47. Lai MC, Kassee C, Besney R, Bonato S, Hull L, Mandy W, et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. The Lancet Psychiatry. 2019;6(10):819–29.

48. Hirvikoski T, Mittendorfer-Rutz E, Boman M, Larsson H, Lichtenstein P, Bölte S. Premature mortality in autism spectrum disorder. Br J Psychiatry. 2016;208(3):232–8.

49. Wiśniowiecka-Kowalnik B, Nowakowska BA. Genetics and epigenetics of autism spectrum disorder—current evidence in the field. J Appl Genet. 2019;60(1):37–47.

50. Kim JY, Son MJ, Son CY, Radua J, Eisenhut M, Gressier F, et al. Environmental risk factors and biomarkers for autism spectrum disorder: an umbrella review of the evidence. Lancet Psychiatry. 2019;6(7):590–600.

51. Wu S, Wu F, Ding Y, Hou J, Bi J, Zhang Z. Advanced parental age and autism risk in children: a systematic review and meta-analysis. Acta Psychiatr Scand. 2017;135(1):29–41.

52. Modabbernia A, Velthorst E, Reichenberg A. Environmental risk factors for autism: an evidence-based review of systematic reviews and meta-analyses. Mol Autism. 2017;8(1):1–16.

53. Elsabbagh M. Linking risk factors and outcomes in autism spectrum disorder: Is there evidence for resilience? BMJ. 2020;368.

54. Fujiwara T, Morisaki N, Honda Y, Sampei M, Tani Y. Chemicals, nutrition, and autism spectrum disorder: A mini-review. Front Neurosci. 2016;10(174):1–7.

55. Davidson M. Vaccination as a cause of autism-myths and controversies. Dialogues Clin Neurosci. 2017;19(4):403–7.

56. Hyman SL, Levy SE, Myers SM. Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. Pediatrics. 2020;145(1).

57. Baumer N, Spence SJ. Evaluation and Management of the Child With Autism Spectrum Disorder. Child Neurol. 2018;24(1):248–75.

58. CDC. Treatment and Intervention Services for Autism Spectrum Disorder [Internet]. https://www.cdc.gov/ncbddd/autism/treatment.html

59. Myers SM, Johnson CP, Lipkin PH, Cartwright JD, Desch LW, Duby JC, et al. Management of children with autism spectrum disorders. Pediatrics. 2007;120(5):1162–82.

60. Weir E, Allison C, Baron-Cohen S. Autism in children: improving screening, diagnosis and support. Prescriber. 2020;31(1):20–4.

61. Schubart JR, Camacho F, Leslie D. Psychotropic medication trends among children and adolescents with autism spectrum disorder in the Medicaid program. Autism. 2014;18(6):631–7.

62. Fan K wai. The National Center for Complementary and Alternative Medicine. Acad Med. 2007;82(10):967.

63. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry. 2016;57(5):585–95.

64. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The Heritability of Autism Spectrum Disorder. JAMA. 2017;318(12):1182–4.

65. Yip BHK, Bai D, Mahjani B, Klei L, Pawitan Y, Hultman CM, et al. Heritable variation, with little or no maternal effect, accounts for recurrence risk to Autism Spectrum Disorder in Sweden. Biol Psychiatry. 2018;83(7):589–97.

66. Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, et al. Association of Genetic and Environmental Factors with Autism in a 5-Country Cohort. JAMA Psychiatry. 2019;76(10):1035–43.

67. Hansen SN, Schendel DE, Francis RW, Windham GC, Bresnahan M, Levine SZ, et al. Recurrence Risk of Autism in Siblings and Cousins: A

Multinational, Population-Based Study. J Am Acad Child Adolesc Psychiatry. 2019;58(9):866–75.

68. Risch N, Hoffmann T, Anderson M, Croen LA, Crether J, Windham G. Familial recurrence of autism spectrum disorder: Evaluating genetic and environmental contributions. Am J Psychiatry. 2014;171(11):1206–13.

69. 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;128(3):e488.

70. 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;8(1):73–81.

71. Iakoucheva LM, Muotri AR, Sebat J. Getting to the Cores of Autism. Cell. 2019;178(6):1287–98.

72. Toma C. Genetic Variation across Phenotypic Severity of Autism. Trends Genet. 2020;36(4):228–31.

73. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Wigler M. The contribution of de novo coding mutations to autism spectrum disorder. Nature. 2014;515(7526):216–21.

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

75. Guo H, Duyzend MH, Coe BP, Baker C, Hoekzema K, Gerdts J, et al. Genome sequencing identifies multiple deleterious variants in autism patients with more severe phenotypes. Genet Med. 2018;21(7):1611–20.

76. Chaste P, Roeder K, Devlin B. The Yin and Yang of Autism Genetics: How Rare De Novo and Common Variations Affect Liability. Annu Rev Genom Hum Genet. 2017;18(921):1–9.

77. Woodbury-Smith M, Scherer SW. Progress in the genetics of autism spectrum disorder. Dev Med Child Neurol. 2018;60(5):445–51.

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

79. Fernandez BA, Scherer SW. Syndromic autism spectrum disorders: moving from a clinically defined to a molecularly defined approach. Dialogues Clin Neurosci. 2017;19(4):353–71.

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

81. Kumar RA, Christian SL. Genetics of Autism Spectrum Disorders. Curr Neurol Neurosci Rep. 2009;9(3):188–97.

82. 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;316(5823):445–9.

83. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, et al. Structural Variation of Chromosomes in Autism Spectrum Disorder. Am J Hum Genet. 2008;82(2):477–88.

84. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Devlin B, State MW. Multiple recurrent de novo copy number variations (CNVs), including duplications of the 7q11.23 Williams-Beuren syndrome region, are strongly associated with autism. Neuron. 2011;70(5):863–85.

85. Levy D, Ronemus M, Yamrom B, Lee Y ha, Leotta A, Kendall J, et al. Rare De Novo and Transmitted Copy-Number Variation in Autistic Spectrum Disorders. Neuron. 2011;70(5):886–97.

86. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, et al. Functional impact of global rare copy number variation in autism spectrum disorders. Nature. 2010;466(7304):368–72.

87. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, Karamohamed S, et al. Novel Submicroscopic chromosomal abnormalities detected in Autism Spectrum Disorder. Biol Psychiatry. 2008;63(12):1111–7.

88. 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;18(10):1795–804.

89. Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. Genet Med. 2013;15(6):478–81.

90. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between Microdeletion and Microduplication at 16p11.2 and Autism. N Engl J Med. 2008;358(7):667–75.

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

92. Vorstman JAS, Parr JR, Moreno-De-Luca D, Anney RJL, Nurnberger Jr JI, Hallmayer JF. Autism genetics: opportunities and challenges for clinical translation. Nat Rev Genet. 2017;18(6):362–76.

93. Butler MG, Rafi SK, Manzardo AM. High-resolution chromosome ideogram representation of currently recognized genes for Autism Spectrum Disorders. Vol. 16, International Journal of Molecular Sciences. 2015. 6464–6495 p.

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

95. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature. 2012;485(237–241).

96. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De Novo Gene Disruptions in Children on the Autistic Spectrum. 2012;74(2):285–99.

97. Du X, Gao X, Liu X, Shen L, Wang K, Fan Y, et al. Genetic Diagnostic Evaluation of Trio-Based Whole Exome Sequencing Among Children With Diagnosed or Suspected Autism Spectrum Disorder. Front Genet. 2018;9(594):1–8.

98. 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. 2012;485(7397):246–50.

99. 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. 2013;485(7397):242–5.

100. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of Autism. Cell. 2020;180:1–17.

101. Hull L, Petrides K V., Mandy W. The Female Autism Phenotype and Camouflaging: a Narrative Review. Rev J Autism Dev Disord. 2020;7(4):306–17.

102. 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. Dendrites Dev Dis. 2012;338(6114):1619–22.

103. Stessman HAF, Xiong B, Coe BP, Wang T, Hoekzema K, Fenckova M, et al. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nat Genet. 2017;49(4):515–26.

104. Guo H, Wang T, Wu H, Long M, Coe BP, Li H, et al. Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. Mol Autism. 2018;9(64):1–12.

105. 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;515(7526):209–15.

106. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. J Autism Dev

Disord. 2015;47(6):582–8.

107. 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;21(2):185–91.

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

109. Ruzzo EK, Pérez-Cano L, Jung JY, Wang L kai, Kashef-Haghighi D, Hartl C, et al. Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. Cell. 2019;178(4):850–66.

110. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variants. Nat Genet. 2014;46(8):881–5.

111. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51(3):431–44.

112. 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.

113. Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB. A noncoding RNA antisense to moesin at 5p14.1 in autism. Sci Transl Med. 2012;4(128).

114. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism. 2017;8(21):1–17.

115. Campbell IM, Shaw CA, Stankiewicz P, Lupski JR. Somatic mosaicism: Implications for disease and transmission genetics. Trends Genet. 2015;31(7):382–92.

116. Poduri A, Evrony GD, Cai X, Walsh CA. Somatic mutation, genomic variation, and neurological disease. Science. 2013;341(6141):1–16.

117. D'Gama AM, Pochareddy S, Li M, Jamuar SS, Reiff RE, Lam ATN, et al. Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms. Neuron. 2015;88(5):910–7.

118. Freed D, Pevsner J. The Contribution of Mosaic Variants to Autism Spectrum Disorder. PLoS Genet. 2016;12(9):1–20.

119. Lim ET, Uddin M, De Rubeis S, Chan Y, Kamumbu AS, Zhang X, et al. Rates, distribution and implications of postzygotic mosaic mutations in autism

spectrum disorder. Nat Neurosci. 2017;20:1217–24.

120. Krupp DR, Barnard RA, Duffourd Y, Evans SA, Mulqueen RM, Bernier R, et al. Exonic Mosaic Mutations Contribute Risk for Autism Spectrum Disorder. Am J Hum Genet. 2017;101:369–90.

121. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409(6846):860–921.

122. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. Segmental Duplications: Organization and Impact Within the Current Human Genome Project Assembly. Genome Res. 2001;11(6):1005–17.

123. Numanagić I, Gökkaya AS, Zhang L, Berger B, Alkan C, Hach F. Fast characterization of segmental duplications in genome assemblies. Bioinformatics. 2018;34(17):i706–14.

124. Zhang L, Lu HHS, Chung WY, Yang J, Li WH. Patterns of segmental duplication in the human genome. Mol Biol Evol. 2005;22(1):135–41.

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

126. Lupski JR. Charcot-Marie-Tooth disease: Lessons in genetic mechanisms. Mol Med. 1998;4(1):3–11.

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

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

129. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte R V., Schwartz S, et al. Recent segmental duplications in the human genome. Science (80-). 2002;297(5583):1003–7.

130. Dennis MY, Eichler EE. Human adaptation and evolution by segmental duplication. Curr Opin Genet Dev. 2016;41:44–52.

131. Kirkpatrick M. How and why chromosome inversions evolve. PLoS Biol. 2010;8(9):1–5.

132. Puig M, Casillas S, Villatoro S, Cáceres M. Human inversions and their functional consequences. Brief Funct Genomics. 2015;14(5):369–79.

133. Escaramís G, Docampo E, Rabionet R. A decade of structural variants: description, history and methods to detect structural variation. Brief Funct Genomics. 2015;14(5):305–14.

134. Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tönnesen T, Carlberg BM, et al. Inversion of the IDS gene resulting from recombination with IDS-

related sequences in a common cause of the Hunter syndrome. Hum Mol Genet. 1995;4(4):615–21.

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

136. Cukier HN, Skaar DA, Rayner-Evans MY, Konidari I, Whitehead PL, Jaworski JM, et al. Identification of Chromosome 7 Inversion Breakpoints in an Autistic Family Narrows Candidate Region for Autism Susceptibility. 2009;2(5):258–66.

137. González JR, Cáceres A, Esko T, Cuscó I, Puig M, Esnaola M, et al. A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. Am J Hum Genet. 2014;94(3):361–72.

138. González JR, Ruiz-Arenas C, Cáceres A, Morán I, López-Sánchez M, Alonso L, et al. Polymorphic inversions underlie the shared genetic susceptibility of obesity-related diseases. Am J Hum Genet. 2020;106(6):846–58.

139. Antonacci F, Kidd JM, Marques-Bonet T, Ventura M, Siswara P, Jiang Z, et al. Characterization of six human disease-associated inversion polymorphisms. Hum Mol Genet. 2009;18(14):2555–66.

140. Caceres A. inveRsion: Inversions in genotype data. R package version 1.38.0. [Internet]. 2020.

141. Cáceres A, González JR. Following the footprints of polymorphic inversions on SNP data: from detection to association tests. Nucleic Acids Res. 2015;43(8):e53:1-11.

142. Puig M, Castellano D, Pantano L, Giner-Delgado C, Izquierdo D, Gayà-Vidal M, et al. Functional Impact and Evolution of a Novel Human Polymorphic Inversion That Disrupts a Gene and Creates a Fusion Transcript. PLoS Genet. 2015;11(10):1–28.

143. Aguado C, Gayà-Vidal M, Villatoro S, Oliva M, Izquierdo D, Giner-Delgado C, et al. Validation and Genotyping of Multiple Human Polymorphic Inversions Mediated by Inverted Repeats Reveals a High Degree of Recurrence. PLoS Genet. 2014;10(3):14–22.

144. Salm MPA, Horswell SD, Hutchison CE, Speedy HE, Yang X, Liang L, et al. The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. Genome Res. 2012;22:1144–53.

145. Cáceres A, González JR. invClust: inference of polymorphic inversions from multivariate analysis on SNP data [Internet]. 2015.

146. Cáceres A, Sindi SS, Raphael BJ, Cáceres M, González JR. Identification of polymorphic inversions from genotypes. BMC Bioinformatics. 2012.

147. Ruiz-Arenas C, Cáceres A, López-Sánchez M, Tolosana I, Pérez-Jurado L, González JR. scoreInvHap: Inversion genotyping for genome-wide association studies. PLoS Genet. 2019;15(7):1–16.

148. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, et al. A common inversion under selection in Europeans. Nat Genet. 2005;37(2):129–37.

149. Baker M, Litvan I, Houlden H, Adamson J, Dickson D, Perez-Tur J, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum Mol Genet. 1999;8(4):711–5.

150. Rao PN, Li W, Vissers LELM, Veltman JA, Ophoff RA. Recurrent inversion events at 17q21.31 microdeletion locus are linked to the MAPT H2 haplotype. Cytogenet Genome Res. 2010;129(4):275–9.

151. Zody MC, Jiang Z, Fung H, Antonacci F, Hillier LW, Cardone MF, et al. Evolutionary Toggling of the MAPT 17q21.31 Inversion Region. Nat Genet. 2008;40(9):1076–83.

152. de Jong S, Chepelev I, Janson E, Strengman E, van den Berg LH, Veldink JH, et al. Common inversion polymorphism at 17q21.31 affects expression of multiple genes in tissue-specific manner. BMC Genomics. 2012;13(458):1–6.

153. Houlden H, Baker M, Morris HR, MacDonald N, Pickering-Brown S, Adamson J, et al. Corticobasal degeneration and progressive supranuclear palsy share a common tau haplotype. Neurology. 2001;56(12):1702–6.

154. Skipper L, Wilkes K, Toft M, Baker M, Lincoln S, Hulihan M, et al. Linkage disequilibrium and association of MAPT H1 in Parkinson disease. Am J Hum Genet. 2004;75(4):669–77.

155. Myers AJ, Kaleem M, Marlowe L, Pittman AM, Lees AJ, Fung HC, et al. The H1c haplotype at the MAPT locus is associated with Alzheimer's disease. Hum Mol Genet. 2005;14(16):2399–404.

156. Sánchez-Juan P, Moreno S, de Rojas I, Hernández I, Valero S, Alegret M, et al. The MAPT H1 Haplotype Is a Risk Factor for Alzheimer's Disease in APOE ε4 Non-carriers. Front Aging Neurosci. 2019;11(327):1–9.

157. Koolen DA, Vissers LELM, Pfundt R, De Leeuw N, Knight SJL, Regan R, et al. A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet. 2006;38(9):999–1001.

158. Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, et al. Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. Nat Genet.

2006;38(9):1032–7.

159. Sugawara H, Harada N, Ida T, Ishida T, Ledbetter DH, Yoshiura KI, et al. Complex low-copy repeats associated with a common polymorphic inversion at human chromosome 8p23. Genomics. 2003;82(2):238–44.

160. Bosch N, Morell M, Ponsa I, Mercader JM, Armengol L, Estivill X. Nucleotide, cytogenetic and expression impact of the human chromosome 8p23.1 inversion polymorphism. PLoS One. 2009;4(12):e8269.

161. Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, Clinical, and Population Relevance of 95 Loci for Blood Lipids. Nature. 2010;466(7307):707–13.

162. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, et al. Association of Systemic Lupus Erythematosus with C8orf13-BLK and ITGAM-ITGAX. N Engl J Med. 2008;358(9):900–9.

163. Ito I, Kawaguchi Y, Kawasaki A, Hasegawa M, Ohashi J, Kawamoto M, et al. Association of the FAM167A-BLK region with systemic sclerosis. Arthritis Rheum. 2010;62(3):890–5.

164. Zhou Y, Li X, Wang G, Li X. Association of FAM167A-BLK rs2736340 Polymorphism with Susceptibility to Autoimmune Diseases: A Meta-Analysis. Immunol Invest. 2016;45(4):336–48.

165. Song GG, Lee YH. Association between BLK polymorphisms and susceptibility to SLE. Z Rheumatol. 2017;76(2):176–82.

166. Namjou B, Ni Y, Harley ITW, Chepelev I, Cobb B, Kottyan LC, et al. The effect of inversion at 8p23 on BLK association with Lupus in Caucasian population. PLoS One. 2014;9(12):1–13.

167. Demirci FY, Wang X, Morris DL, Feingold E, Bernatsky S, Pineau C, et al. Multiple signals at the extended 8p23 locus are associated with susceptibility to systemic lupus erythematosus (SLE). J Med Genet. 2017;54(6):381–9.

168. Hollox EJ, Barber JCK, Brookes AJ, Armour JAL. Defensins and the dynamic genome: What we can learn from structural variation at human chromosome band 8p23.1. Genome Res. 2008;18(11):1686–97.

169. Giglio S, Calvari V, Gregato G, Gimelli G, Camanini S, Giorda R, et al. Heterozygous submicroscopic inversions involving olfactory receptor-gene clusters mediate the recurrent t(4;8)(p16;p23) translocation. Am J Hum Genet. 2002;71(2):276–85.

170. Hillier LDW, Fulton RS, Fulton LA, Graves TA, Pepin KH, Wagner-McPherson C, et al. The DNA sequence of human chromosome 7. Nature. 2003;424(6945):157–64.

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

172. 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;69(1):1–13.

173. 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;66(1):47–68.

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

175. Williams J, Barratt-Boyes B, Lowe M. Supravalvular aortic stenosis. Circulation. 1961;XXIV:1311–8.

176. Beuren A, Apitz J, Harmjanz D. Supravalvular aortic stenosis in association with mental retardation and a certain facial appearance. Circulation. 1962;26:1235–40.

177. 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;43(3):266–73.

178. 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;47(5):312–20.

179. Schubert C. The genomic basis of the Williams - Beuren syndrome. Cell Mol Life Sci. 2009;66(7):1178–97.

180. 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 Part C Semin Med Genet. 2010;154C(2):220–8.

181. Pober B. Williams-Beuren Syndrome. N Engl J Med. 2010;362:239–52.

182. Morris CA, Demsey SA, Leonard CO, Dilts C, Blackburn BL. Natural history of Williams syndrome: Physical characteristics. J Pediatr. 1988;113(2):318–26.

183. Pober BR, Johnson M, Urban Z. Mechanisms and treatment of cardiovascular disease in Williams-Beuren syndrome. J Clin Invest. 2008;118(5):1606–15.

184. 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. 2004;127A(3):234–7.

185. 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. 2004;131A(3):255–64.

186. Mervis CB, Klein-tasman BP. Williams Syndrome: Cognition, personality and adaptive behavior. MRDD Res Rev. 2000;158(6):148–58.

187. Doyle TF, Bellugi U, Korenberg JR, Graham J. "Everybody in the world is my friend" Hypersociability in young children with Williams syndrome. Am J Med Genet. 2004;124A(3):263–73.

188. Dykens EM. Anxiety, fears, and phobias in persons with Williams syndrome. Dev Neuropsychol. 2003;23(1–2):291–316.

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

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

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

192. Herguner S, Motavalli Mukaddes N. Autism and Williams syndrome: A case report. World J Biol Psychiatry. 2006;7(3):186–8.

193. 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.

194. Klein-Tasman BP, Mervis CB. Autism Spectrum Symptomatology among Children with Duplication 7q11.23 syndrome. J Autism Dev Disord. 2018;48(6):1982–94.

195. Masson J, Demily C, Chatron N, Labalme A, Rollat-Farnier P-A, Schluth-Bolard C, et al. Molecular investigation, using chromosomal microarray and whole exome sequencing, of six patients affected by Williams Beuren syndrome and Autism Spectrum Disorder. Orphanet J Rare Dis. 2019;14(121):1–6.

196. 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. Lancet Psychiatry. 2015;2:1–8.

197. 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;13(5):444–67.

198. Somerville MJ, Mervis CB, Young EJ, Seo EJ, Del Campo M, Bamforth S, et al. Severe expressive-language delay related to duplication of the Williams-Beuren locus. N Engl J Med. 2005;353(16):1694–701.

199. Earhart BA, Williams ME, Zamora I, Randolph LM, Votava-Smith JK, Marcy SN. Phenotype of 7q11.23 duplication: A family clinical series. Am J Med Genet Part A. 2016;9999A:1–6.

200. 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. 2009;52:94–100.

201. 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;18(3):108–16.

202. 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;47:271–5.

203. 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;167A(12):2916–35.

204. Abbas E, Cox D, Smith T, Butler M. The 7q11.23 Microduplication Syndrome: A Clinical Report with Review of Literature. J Pediatr Genet. 2016;5:129–40.

205. 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;167(7):1436–50.

206. Mulle JG, Pulver AE, McGrath JM, Wolyniec P, 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;75(5):371–7.

207. Geschwind DH. Autism: Many Genes, Common Pathways? Cell. 2008;135(3):391–5.

208. Rylaarsdam L, Guemez-Gamboa A. Genetic Causes and Modifiers of Autism Spectrum Disorder. Front Cell Neurosci. 2019;13(385):1–15.

209. Guang S, Pang N, Deng X, Yang L, He F, Wu L, et al. Synaptopathology involved in autism spectrum disorder. Front Cell Neurosci. 2018;12(470):1–16.

210. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet. 2014;94(5):677–94.

211. Courchesne E, Pramparo T, Gazestani VH, Lombardo M V., Pierce K, Lewis NE. The ASD Living Biology: from cell proliferation to clinical phenotype. Mol Psychiatry. 2019;24(1):88–107.

212. Quesnel-Vallières M, Weatheritt RJ, Cordes SP, Blencowe BJ. Autism spectrum disorder: insights into convergent mechanisms from transcriptomics. Nat Rev Genet. 2019;20(1):51–63.

213. Parikshak NN, Gandal MJ, Geschwind DH, Angeles L, Angeles L, Angeles L. Systems biology and gene networks in neurodevelopm. Nat Rev Genet. 2015;16(8):441–58.

214. Purcell AE, Jeon OH, Zimmerman AW, Blue ME, Pevsner J. Postmortem brain abnormalities of the glutamate neurotransmitter system in autism. Neurology. 2001;57(9):1618–28.

215. Rubenstein J, Merzenich M. Model of autism: increased ratio of excitation/inhibition in key neural systems. Genes Brain Behav. 2003;2(5):255–67.

216. Voineagu I, Wang X, Johnston P, Lowe JK, Tian Y, Mill J, et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. 2013;474(7351):380–4.

217. 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.

218. 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):1–8.

219. Parikshak NN, Swarup V, Belgard TG, Irimia M, Gandal MJ, Hartl C, et al. Genome-wide changes in lncRNA, splicing and regional gene expression patterns in autism. Nature. 2016;540(7633):423–7.

220. Derosa BA, El Hokayem J, Artimovich E, Garcia-Serje C, Phillips AW, Van Booven D, et al. Convergent Pathways in Idiopathic Autism Revealed by Time Course Transcriptomic Analysis of Patient-Derived Neurons. Sci

Rep. 2018;8(8423):1–15.

221. Tylee DS, Hess JL, Quinn TP, Barve R, Huang H, Zhang-James Y, et al. Blood transcriptomic comparison of individuals with and without autism spectrum disorder: A combined-samples mega-analysis. Am J Med Genet Part B Neuropsychiatr Genet. 2017;174(3):181–201.

222. Ziats MN, Rennert OM. Aberrant Expression of Long Noncoding RNAs in Autistic Brain. J Mol Neurosci. 2013;49(3):589–93.

223. Wang Y, Zhao X, Ju W, Flory M, Zhong J, Jiang S, et al. Genome-wide differential expression of synaptic long noncoding RNAs in autism spectrum disorder. Transl Psychiatry. 2015;5(10):e660.

224. Tang J, Yu Y, Yang W. Long noncoding RNA and its contribution to autism spectrum disorders. CNS Neurosci Ther. 2017;23(8):645–56.

225. 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. 2012;485:246–50.

226. Girirajan S, Dennis MY, Baker C, Malig M, Coe BP, Campbell CD, et al. Refinement and Discovery of New Hotspots of Copy-Number Variation Associated with Autism Spectrum Disorder. Am J Hum Genet. 2013;92:221–37.

227. 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.

228. Consortium EA. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536(7616):285–91.

229. Kopp ND, Parrish PCR, Lugo M, Dougherty JD, Kozel BA. Exome sequencing of 85 Williams-Beuren syndrome cases rules out coding variation as a major contributor to remaining variance in social behavior. Mol Genet Genomic Med. 2018;6(5):749-765.

230. Hansen J, Snow C, Tuttle E, Ghoneim DH, Yang CS, Spencer A, et al. De novo mutations in SIK1 cause a spectrum of developmental epilepsies. Am J Hum Genet. 2015;96(4):682–90.

231. 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. Shamir R, editor. PLoS Comput Biol. 2011;7(1):e1001054.

232. Khattak S, Brimble E, Zhang W, Zaslavsky K, Strong E, Ross PJ, et al. Human induced pluripotent stem cell derived neurons as a model for Williams-Beuren syndrome. Mol Brain. 2015;8(1):1–11.

233. Lalli MA, Jang J, Park JHC, Wang Y, Guzman E, Zhou H, et al. Haploinsufficiency of BAZ1B contributes to Williams syndrome through transcriptional dysregulation of neurodevelopmental pathways. Hum Mol Genet. 2016;25(7):1294–306.

234. Chailangkarn T, Trujillo CA, Freitas BC, Hrvoj-mihic B, Herai RH, Yu DX, et al. A human neurodevelopmental model for Williams syndrome. Nature. 2016;536(7616):338–43.

235. 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;47(2):132–41.

236. Zanella M, Vitriolo A, Andirko A, Martins PT, Sturm S, O'Rourke T, et al. Dosage analysis of the 7q11.23 Williams region identifies BAZ1B as a major human gene patterning the modern human face and underlying self-domestication. Sci Adv. 2019;5(12):eaaw7908.

237. Vuillaume ML, Cogné B, Jeanne M, Boland A, Ung DC, Quinquis D, et al. Whole genome sequencing identifies a de novo 2.1 Mb balanced paracentric inversion disrupting FOXP1 and leading to severe intellectual disability. Clin Chim Acta. 2018;485:218–23.

238. Pain O, Pocklington AJ, Holmans PA, Bray NJ, O'Brien HE, Hall LS, et al. Novel Insight Into the Etiology of Autism Spectrum Disorder Gained by Integrating Expression Data With Genome-wide Association Statistics. Biol Psychiatry. 2019;86(4):265–73.

239. Liu T, Porter J, Zhao C, Zhu H, Wang N, Sun Z, et al. TADKB: Family classification and a knowledge base of topologically associating domains. BMC Genomics. 2019;20(1):217.

240. Maass PG, Barutcu AR, Rinn JL. Interchromosomal interactions: A genomic love story of kissing chromosomes. J Cell Biol. 2019;218(1):27–38.

241. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat Rev Genet. 2018;19(7):453–67.

242. Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, et al. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet. 2017;49(1):36–45.

243. Brandler WM, Antaki D, Gujral M, Kleiber ML, Whitney J, Maile MS, et al. Paternally inherited cis-regulatory structural variants are associated with autism. Science. 2018;360(6386):327–31.

244. Branco MR, Pombo A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent

associations. PLoS Biol. 2006;4(5):780–8.

245. Kaufmann S, Fuchs C, Gonik M, Khrameeva EE, Mironov AA, Frishman DF. Inter-chromosomal contact networks provide insights into mammalian chromatin organization. PLoS One. 2015;10(5):1–25.

246. Loviglio MN, Leleu M, Männik K, Passeggeri M, Giannuzzi G, Van Der Werf I, et al. Chromosomal contacts connect loci associated with autism, BMI and head circumference phenotypes. Mol Psychiatry. 2017;22(6):836–49.

247. D'Souza-Schorey C, Chavrier P. ARF proteins: Roles in membrane traffic and beyond. Nat Rev Mol Cell Biol. 2006;7(5):347–58.

248. Jaworski J. ARF6 in the nervous system. Eur J Cell Biol. 2007 Sep 14;86(9):513–24.

249. Rogers EJ, Jada R, Scharagenheim-Rozales K, Sah M, Cortes M, Florence M, et al. An IQSEC2 Mutation Associated With Intellectual Disability and Autism Results in Decreased Surface AMPA Receptors. Front Mol Neurosci. 2019;12(43):1–18.

250. Wang H. Lipid rafts: a signaling platform linking cholesterol metabolism to synaptic deficits in autism spectrum disorders. Front Behav Neurosci. 2014;8(104):1–6.

251. Dziobek I, Gold SM, Wolf OT, Convit A. Hypercholesterolemia in Asperger syndrome: Independence from lifestyle, obsessive-compulsive behavior, and social anxiety. Psychiatry Res. 2007;149(1–3):321–4.

252. Tint GS, Irons M, Elias ER, Batta AK, Frieden R, Chen TS, et al. Defective Cholesterol Biosynthesis Associated with the Smith-Lemli-Opitz Syndrome. N Engl J Med. 1994;330(2):107–13.

253. 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;52(4):248–55.

254. Rojas DC. The role of glutamate and its receptors in autism and the use of glutamate receptor antagonists in treatment. J Neural Transm. 2014;121(8):891–905.

255. Choudhury PR, Lahiri S, Rajamma U. Glutamate mediated signaling in the pathophysiology of autism spectrum disorders. Pharmacol Biochem Behav. 2012;100(4):841–9.

256. Tarabeux J, Kebir O, Gauthier J, Hamdan FF, Xiong L, Piton A, et al. Rare mutations in N-methyl-D-aspartate glutamate receptors in autism spectrum disorders and schizophrenia. Transl Psychiatry. 2011;1:1–7.

257. Gai X, Xie H, Perin J, Takahashi N, Murphy K, Wenocur A, et al. Rare

structural variation of synapse and neurotransmission genes in autism. Mol Psychiatry. 2012;17:402–11.

258. Sekino Y, Kojima N, Shirao T. Role of actin cytoskeleton in dendritic spine morphogenesis. Neurochem Int. 2007;51(2–4):92–104.

259. Griesi-Oliveira K, Suzuki AM, Alves AY, Mafra ACCN, Yamamoto GL, Ezquina S, et al. Actin cytoskeleton dynamics in stem cells from autistic individuals. Sci Rep. 2018;8(1):1–10.

260. Duffney L, Zhong P, Wei J, Matas E, Cheng J, Qin L, et al. Autism-like deficits in Shank3-deficient mice are rescued by targeting actin regulators. Cell Rep. 2015;11(9):1400–13.

261. Joensuu M, Lanoue V, Hotulainen P. Dendritic spine actin cytoskeleton in autism spectrum disorder. Prog Neuro-Psychopharmacology Biol Psychiatry. 2018;84(Pt B):362–81.

262. Hu VW, Frank BC, Heine S, Lee NH, Quackenbush J. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. BMC Genomics. 2006;7:118.

263. Gregg J, Lit L, Baron C, Hertz-Picciotto I, Walker W, Davis R, et al. Gene expression changes in children with autism. Genomics. 2008;91:22–9.

264. Enstrom A, Lit L, Onore C, Gregg J, Hansen R, Pessah I, et al. Altered gene expression and function of peripheral blood natural killer cells in children with autism. Brain Behav Immun. 2009;23(1):124–33.

265. Chien WH, Gau SSF, Chen CH, Tsai WC, Wu YY, Chen PH, et al. Increased gene expression of FOXP1 in patients with autism spectrum disorders. Mol Autism. 2013;4(1):1–8.

266. Kong SW, Collins CD, Shimizu-Motohashi Y, Holm IA, Campbell MG, Lee IH, et al. Characteristics and Predictive Value of Blood Transcriptome Signature in Males with Autism Spectrum Disorders. PLoS One. 2012;7(12):e49475.

267. Pramparo T, Pierce K, Lombardo M, Carter Barnes C, Marinero S, Ahrens-Barbeau C, et al. Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. JAMA Psychiatry. 2015;72(4):386–94.

268. Filosi M, Kam-Thong T, Essioux L, Muglia P, Trabetti E, Spooren W, et al. Transcriptome signatures from discordant sibling pairs reveal changes in peripheral blood immune cell composition in Autism Spectrum Disorder. Transl Psychiatry. 2020;10(1):106.

269. Bjørklund G, Saad K, Chirumbolo S, Kern JK, Geier DA, Geier MR, et al. Immune dysfunction and neuroinflammation in autism spectrum

disorder. Acta Neurobiol Exp (Wars). 2016;76(4):257–68.

270. Siniscalco D, Schultz S, Brigida AL, Antonucci N. Inflammation and Neuro-Immune Dysregulations in Autism Spectrum Disorders. Pharm. 2018;11(56):1–14.

271. Pangrazzi L, Balasco L, Bozzi Y. Molecular Sciences Oxidative Stress and Immune System Dysfunction in Autism Spectrum Disorders. Int J Mol Sci. 2020;21(3288):1–14.

272. Voineagu I, Eapen V. Converging pathways in autism spectrum disorders: Interplay between synaptic dysfunction and immune responses. Front Hum Neurosci. 2013;7:738.

273. Rodriguez JI, Kern JK. Evidence of microglial activation in autism and its possible role in brain underconnectivity. Neuron Glia Biol. 2011;7(2–4):205–13.

274. Comi AM, Zimmerman AW, Frye VH, Law PA, Peeden JN. Familial clustering of autoimmune disorders and evaluation of medical risk factors in autism. J Child Neurol. 1999;14(6):388–94.

275. Atladóttir HÓ, Pedersen MG, Thorsen P, Mortensen PB, Deleuran B, Eaton WW, et al. Association of family history of autoimmune diseases and autism spectrum disorders. Pediatrics. 2009;124(2):687–94.

276. Wu S, Ding Y, Wu F, Li R, Xie G, Hou J, et al. Family history of autoimmune diseases is associated with an increased risk of autism in children: A systematic review and meta-analysis. Neurosci Biobehav Rev. 2015;55:322–32.

277. Rom AL, Wu C Sen, Olsen J, Jawaheer D, Hetland ML, Mørch LS. Parental Rheumatoid Arthritis and Autism Spectrum Disorders in Offspring: A Danish Nationwide Cohort Study. J Am Acad Child Adolesc Psychiatry. 2018;57(1):28-32.e1.

278. Zhu Z, Tang S, Deng X, Wang Y. Maternal Systemic Lupus Erythematosus, Rheumatoid Arthritis, and Risk for Autism Spectrum Disorders in Offspring: A Meta-analysis. J Autism Dev Disord. 2020;50(8):2852–9.

279. Braunschweig D, Krakowiak P, Duncanson P, Boyce R, Hansen RL, Ashwood P, et al. Autism-specific maternal autoantibodies recognize critical proteins in developing brain. Transl Psychiatry. 2013;3(7):e277-9.

280. Zimmerman AW, Connors SL, Matteson KJ, Lee LC, Singer HS, Castaneda JA, et al. Maternal antibrain antibodies in autism. Brain Behav Immun. 2007;21(3):351–7.

281. Tsai PH, Yu KH, Chou IJ, Luo SF, Tseng WY, Huang LH, et al. Risk of autism spectrum disorder in children born to mothers with systemic lupus erythematosus and rheumatoid arthritis in Taiwan. Jt Bone Spine.

2018;85(5):599–603.

282. Chen S wei, Zhong X shan, Jiang L na, Zheng X yan, Xiong Y quan, Ma S juan, et al. Maternal autoimmune diseases and the risk of autism spectrum disorders in offspring: A systematic review and meta-analysis. Behav Brain Res. 2016;296:61–9.

283. Kronzer VL, Crowson CS, Sparks JA, Myasoedova E, Davis J. Family history of rheumatologic, autoimmune, and non-autoimmune diseases and risk of rheumatoid arthritis. Arthritis Care Res (Hoboken). 2019;73(2):180–7.

284. Warren RP, Odell JD, Warren WL, Burger RA, Maciulis A, Daniels WW, et al. Strong association of the third hypervariable region of HLA-DR β1 with autism. J Neuroimmunol. 1996;67(2):97–102.

285. Torres AR, Maciulis A, Stubbs EG, Cutler A, Odell D. The transmission disequilibrium test suggests that HLA-DR4 and DR13 are linked to autism spectrum disorder. Hum Immunol. 2002;63(4):311–6.

286. 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. 2006;67(4–5):346–51.

287. Roudier J. Association of MHC and rheumatoid arthritis association of RA with HLA-DR4: The role of repertoire selection. Arthritis Res. 2000;2(3):217–20.

288. Fries JF, Wolfe F, Apple R, Erlich H, Bugawan T, Holmes T, et al. HLA-DRB1 genotype associations in 793 white patients from a rheumatoid arthritis inception cohort: Frequency, severity, and treatment bias. Arthritis Rheum. 2002;46(9):2320–9.

289. Crespi BJ, Thiselton DL. Comparative immunogenetics of autism and schizophrenia. Genes, Brain Behav. 2011;10(7):689–701.

290. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014 Nov 13;515(7526):209–15.

291. Yoo HJ, Cho IH, Park M, Cho E, Cho SC, Kim BN, et al. Association between PTGS2 polymorphism and autism spectrum disorders in Korean trios. Neurosci Res. 2008;62(1):66–9.

292. Willsey AJ, Sanders SJ, Li M, Dong S, Tebbenkamp A, Muhle RA, et al. Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. Cell. 2014;155(5):997–1007.

293. Ashwood P, Krakowiak P, Hertz-Picciotto I, Hansen R, Pessah I, Van de Water J. Elevated plasma cytokines in autism spectrum disorders provide evidence of immune dysfunction and are associated with impaired

behavioral outcome. Brain Behav Immun. 2011 Jan 1;25(1):40–5.

294. Saad K, Abdallah AEM, Abdel-Rahman AA, Al-Atram AA, Abdel-Raheem YF, Gad EF, et al. Polymorphism of interleukin-1β and interleukin-1 receptor antagonist genes in children with autism spectrum disorders. Prog Neuro-Psychopharmacology Biol Psychiatry. 2020;103:109999.

295. Pekkoc Uyanik KC, Kalayci Yigin A, Dogangun B, Seven M. Evaluation of IL1B rs1143634 and IL6 rs1800796 Polymorphisms with Autism Spectrum Disorder in the Turkish Children. Immunol Invest. 2021;1–12.

296. Nagamani SCS, Erez A, Shen J, Li C, Roeder E, Cox S, et al. Clinical spectrum associated with recurrent genomic rearrangements in chromosome 17q12. Eur J Hum Genet. 2010;18(3):278–84.

297. Rasmussen M, Vestergaard EM, Graakjaer J, Petkov Y, Bache I, Fagerberg C, et al. 17q12 deletion and duplication syndrome in Denmark—A clinical cohort of 38 patients and review of the literature. Am J Med Genet Part A. 2016;170(11):2934–42.

298. Yamamoto T, Togawa M, Shimada S, Sangu N, Shimojima K, Okamoto N. Narrowing of the responsible region for severe developmental delay and autistic behaviors in WAGR syndrome down to 1.6Mb including PAX6, WT1, and PRRG4. Am J Med Genet Part A. 2014;164(3):634–8.

299. Zhang L, Qin Z, Ricke KM, Cruz SA, Stewart AFR, Chen HH. Hyperactivated PTP1B phosphatase in parvalbumin neurons alters anterior cingulate inhibitory circuits and induces autism-like behaviors. Nat Commun. 2020;11(1):1017.

300. Krishnan N, Krishnan K, Connors CR, Choy MS, Page R, Peti W, et al. PTP1B inhibition suggests a therapeutic strategy for Rett syndrome. J Clin Invest. 2015;125(8):3163–77.

301. Eulalio A, Mano M, Ferro MD, Zentilin L, Sinagra G, Zacchigna S, et al. Functional screening identifies miRNAs inducing cardiac regeneration. Nature. 2012;492(7429):376–81.

302. Bao MH, Li JM, Zhou QL, Li GY, Zeng J, Zhao J, et al. Effects of miR-590 on oxLDL-induced endothelial cell apoptosis: Roles of p53 and NF-κB. Mol Med Rep. 2016;13(1):867–73.

303. Wang WT, Qi Q, Zhao P, Li CY, Yin XY, Yan RB. miR-590-3p is a novel microRNA which suppresses osteosarcoma progression by targeting SOX9. Biomed Pharmacother. 2018;107:1763–9.

304. Xu BB, Gu ZF, Ma M, Wang JY, Wang HN. MicroRNA-590-5p suppresses the proliferation and invasion of non-small cell lung cancer by regulating GAB1. Eur Rev Med Pharmacol Sci. 2018;22(18):5954–63.

305. Wu X, Li W, Zheng Y. Recent Progress on Relevant microRNAs in Autism

Spectrum Disorders. Int J Mol Sci. 2020;21:5904.

306. Fregeac J, Colleaux L, Nguyen LS. The emerging roles of MicroRNAs in autism spectrum disorders. Neurosci Biobehav Rev. 2016;71:729–38.

307. Hu Y, Ehli EA, Boomsma DI. MicroRNAs as biomarkers for psychiatric disorders with a focus on autism spectrum disorder: Current progress in genetic association studies, expression profiling, and translational research. Autism Res. 2017;10(7):1184–203.

308. Constantin L. The Role of MicroRNAs in Cerebellar Development and Autism Spectrum Disorder During Embryogenesis. Mol Neurobiol. 2017;54(9):6944–59.

309. Bilinovich SM, Lewis K, Grepo N, Campbell DB. The long noncoding RNA RPS10P2-AS1 is implicated in autism spectrum disorder risk and modulates gene expression in human neuronal progenitor cells. Front Genet. 2019;10(970):1–7.

310. Tamizkar KH, Ghafouri-Fard S, Omrani MD, Pouresmaeili F, Arsang-Jang S, Taheri M. Altered expression of lncRNAs in autism spectrum disorder. Metab Brain Dis. 2021;1–8.

311. Khanzada NS, Butler MG, Manzard AM. Geneanalytics pathway analysis and genetic overlap among autism spectrum disorder, bipolar disorder and Schizophrenia. Int J Mol Sci. 2017;18(3):1–11.

312. Yao X, Glessner JT, Li J, Qi X, Hou X, Zhu C, et al. Translational Psychiatry Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. Transl Psychiatry. 2021;11:69.

313. Weingarten LS, Dave H, Li H, Crawford DA. Developmental expression of P5 ATPase mRNA in the mouse. Cell Mol Biol Lett. 2012;17(1):153–70.

314. Anazi S, Maddirevula S, Salpietro V, Asi YT, Alsahli S, Alhashem A, et al. Expanding the genetic heterogeneity of intellectual disability. Hum Genet. 2017;136(11–12):1419–29.

315. Pönniö T, Conneely OM. nor-1 Regulates Hippocampal Axon Guidance, Pyramidal Cell Survival, and Seizure Susceptibility. Mol Cell Biol. 2004;24(20):9070–8.

316. Novak G, Zai CC, Mirkhani M, Shaikh S, Vincent JB, Meltzer H, et al. Replicated association of the NR4A3 gene with smoking behaviour in schizophrenia and in bipolar disorder. Brain Behav. 2010;9(8):910–7.

317. Meyer J, Johannssen K, Freitag CM, Schraut K, Teuber I, Hahner A, et al. Rare variants of the gene encoding the potassium chloride co-transporter 3 are associated with bipolar disorder. Int J Neuropsychopharmacol. 2005;8(4):495–504.

318. Kato T. Molecular genetics of bipolar disorder and depression. Psychiatry Clin Neurosci. 2007;61(1):3–19.

319. Howard HC, Mount DB, Rochefort D, Byun N, Dupré N, Lu J, et al. The K-Cl cotransporter KCC3 is mutant in a severe peripheral neuropathy associated with agenesis of the corpus callosum. Nat Genet. 2002;32(3):384–92.

320. Lee KY, Ahn YM, Joo EJ, Chang JS, Kim YS. The association of DUSP6 gene with schizophrenia and bipolar disorder: Its possible role in the development of bipolar disorder. Mol Psychiatry. 2006;11(5):425–6.

321. Kim SH, Shin SY, Lee KY, Joo EJ, Song JY, Ahn YM, et al. The genetic association of DUSP6 with bipolar disorder and its effect on ERK activity. Prog Neuro-Psychopharmacology Biol Psychiatry. 2012;37(1):41–9.

322. Yoon IS, Li PP, Siu KP, Kennedy JL, Macciardi F, Cooke RG, et al. Altered TRPC7 gene expression in bipolar-I disorder. Biol Psychiatry. 2001;50(8):620–6.

323. Phelan KD, Shwe UT, Abramowitz J, Birnbaumer L, Zheng F. Critical role of canonical transient receptor potential channel 7 in initiation of seizures. Proc Natl Acad Sci U S A. 2014;111(31):11533–8.

ANNEX

I. Sharing the generated knowledge with society

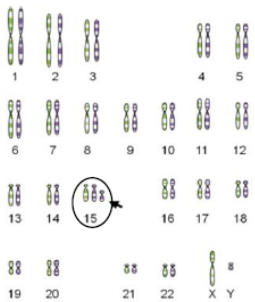
The disclosure of scientific advances to society is of vital importance. This is why, we generated informative leaflets containing general information about ASD, its known genetic causes and different genetic techniques useful to detect them as well as a little insight into genetic counseling. The leaflet shown below was distributed to 6 hospitals and 14 ASD patient associations, mostly located in Catalonia, with a total distribution of 1,330 copies. Also, it was distributed through an online platform from CIBERER (Centro de Investigación Biomédica en Red de Enfermedades Raras), which can be found in the following link:
http://www.ciberer.es/media/919727/asd_triptic_v5-1-1.pdf

ENTITY	COPIES
Rubí TEA	25
Fundació Junts Autisme	80
Aprenem	15
Federació Autisme Catalunya	30
Federació Catalana d'Autisme i Asperger	Online
Associació Síndrome d'Asperger de Catalunya	100
Fundació Autisme Mas Casadavall (FAMC)	70
Associació Obrir-se al món	50
Associació Síndrome d'Asperger d'Osona	100
Associació Asperger del Camp de Tarragona (ASPERCAMP)	100
Associació Síndrome d'Asperger Girona (SAGI)	145
Associació TEA - Asperger Vallés Occidental	15
Asociación Desarrollo Autismo Albacete	80
Federación Autismo Galicia	20
Hospital Vall Hebrón	100
Hospital del Mar	40
Consorci Corporació Sanitària Parc Taulí de Sabadell	190
Hospital Universitario Cruces	20
Dpto. Pediatría, Facultad de Medicina, Univ. de Zaragoza	50
Hospital Materno-Infantil Virgen de la Arrixaca (Murcia)	100

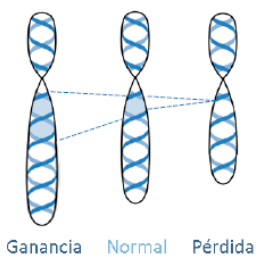
¿QUÉ PRUEBAS GENÉTICAS EXISTEN ACTUALMENTE?

- TEST DEL X-FRÁGIL Se recomienda realizar el test de detección del síndrome X-frágil en varones con TEA dado que un 0,5-6% de ellos presentan dicho síndrome.

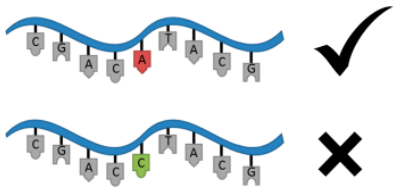
- CARIOTIPO Este test permite detectar las **anomalías cromosómicas de gran tamaño.**



- MICROARRAY Permite determinar si existen **pequeñas ganancias o pérdidas** de material genético.



- SECUENCIACIÓN DEL EXOMA o GENOMA. Permite conocer la **secuencia de nucleótidos** de los genes e identificar cambios en ella. Estas pruebas se encuentran en fase de desarrollo y no se ofrecen de manera rutinaria en los hospitales.



¿CUÁL ES LA FUNCIÓN DE UN ASESOR GENÉTICO?

El **asesor genético** es un profesional de la salud con formación específica en el campo de la genética médica.

Su función es ofrecer **información y apoyo** a las personas que tienen riesgo de padecer o transmitir una **enfermedad genética.** Interpretan la información sobre la enfermedad, analizan los **patrones de herencia,** los **riesgos** de recurrencia u ocurrencia y revisan las diferentes **opciones reproductivas** y de manejo clínico de la familia.

¿DÓNDE PUEDO OBTENER MÁS INFORMACIÓN?

- Sociedad Española de Asesoramiento Genético
- Instituto de Salud Carlos III
- Instituto de Salud Americano
- Unidad de Genética. Universitat Pompeu Fabra
- Área de Genética Clínica y Molecular del Hospital Vall d'Hebrón

ciberer isciii
upf. Universitat Pompeu Fabra Barcelona
PI16/00369-FEDER

1 de cada 100 personas presenta Trastorno del Espectro Autista (TEA)

Actualmente se puede identificar una causa genética en el 30% de las personas con TEA

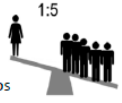
Las personas con TEA y sus familias pueden beneficiarse del asesoramiento genético

¿QUÉ ES EL TRASTORNO DEL ESPECTRO AUTISTA (TEA)?

Se denomina **espectro** debido a que incluye distintas entidades. Los síntomas y la gravedad de los mismos pueden ser muy distintos entre las personas, pero incluyen:

- Dificultades en las interacciones sociales y en la comunicación verbal y no verbal
- Intereses restringidos y desarrollo de actividades de forma repetitiva

¿CON QUÉ FRECUENCIA SE PRESENTA?



- Afecta a 1 de cada 100 personas
- Es 5 veces más frecuente en chicos que en chicas

¿CUÁL ES LA CAUSA DEL TEA?

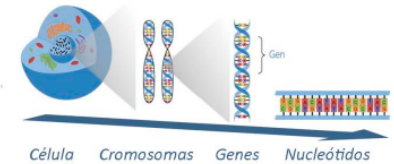
Existen diferentes factores que pueden influir en el riesgo a desarrollar TEA, pero **en la mayoría de los casos no se puede establecer una causa única.** Hay una gran cantidad de **genes**, que en combinación con **factores ambientales**, pueden estar implicados en su desarrollo.

En aproximadamente el **30% de los casos**, la causa principal es una alteración genética.

¿QUÉ SON LOS GENES?

En nuestras **células** se encuentra el **ADN**, una molécula que contiene toda la información que determina nuestras características. Esta información se organiza en **cromosomas**. Disponemos de 23 pares de cromosomas; cada par, está formado por un cromosoma heredado del padre y otro de la madre.

A su vez, cada cromosoma contiene un gran número de **genes** o instrucciones. Los genes están formados por la secuencia de 4 **nucleótidos** (A, C, T, G).



¿QUÉ ALTERACIONES GENÉTICAS DAN LUGAR AL TEA?

En el 30% de los casos en los que se puede identificar una causa genética, ésta puede ser de distintos tipos:

Síndromes asociados

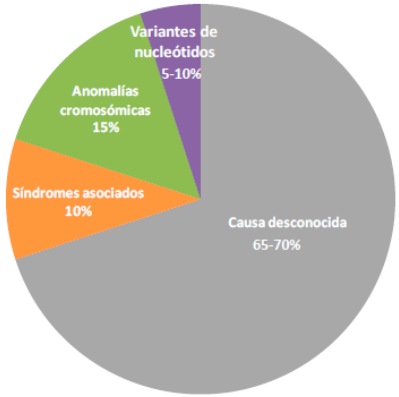
- Existen más de 100 **síndromes** que incluyen el TEA como característica
- Los más frecuentes son el síndrome del X-frágil y la esclerosis tuberosa

Anomalías cromosómicas

- Un fragmento de un **cromosoma** está alterado
- Pueden ser de gran tamaño y afectar a muchos genes, o más pequeñas y alterar pocos genes
- Pueden suponer pérdidas de ADN, ganancias o cambios en su orden
- Las más frecuentes son la ganancia de la región 15q11-q13 del cromosoma 15 y la pérdida de la región 16p11.2 del cromosoma 16

Variantes de nucleótidos

- Cambios en la **secuencia de nucleótidos** de genes asociados al TEA como *CHD8, ARID1B y SCN2A*



Estas alteraciones pueden ser heredadas de progenitores sanos o con manifestaciones leves, o bien aparecer *de novo*. Pueden estar presentes en todas las células del organismo o en determinados tejidos.

¿QUÉ PROBABILIDAD HAY DE TENER UN SEGUNDO HIJO CON TEA?

La respuesta depende de si ha sido posible identificar una causa genética en el primer hijo afectado. Cuando no se puede identificar la causa, se estima que el riesgo para la misma pareja de tener un segundo hijo afectado es entre el **10-27%** en cada nuevo embarazo. Sin embargo, si se conoce la causa es posible establecer el riesgo con mayor exactitud.

Las familias que deseen más información pueden solicitar una visita al **servicio de genética** de un hospital, así como contactar con un **asesor genético**. En esta visita, es probable que se plantee la posibilidad de realizar algún tipo de **prueba genética** para intentar determinar la causa.