

Alma Mater Studiorum – Università di Bologna

**DOTTORATO DI RICERCA IN
BIOLOGIA CELLULARE E MOLECOLARE**

Ciclo 34

Settore Concorsuale: 05/I1 - GENETICA

Settore Scientifico Disciplinare: BIO/18 - GENETICA

Contribution of rare damaging variants in Autism Spectrum Disorder

Presentata da: Marta Viggiano

Coordinatore Dottorato

Prof. Vincenzo Scarlato

Supervisore

Prof.ssa Elena Maestrini

Co-Supervisore

Prof.ssa Elena Bacchelli

Esame finale anno 2022

Abstract

Autism Spectrum Disorder (ASD) is a heterogeneous and highly heritable neurodevelopmental disorder, affecting more than 1% of the population. The core symptoms usually appear within the third year of life and are characterized by impairments in social interactions and communication, repetitive behaviours and restricted interests. ASD is a multifactorial disease, with a complex genetic architecture, consisting of a combination of common low-risk and more penetrant rare variants. Particularly, rare variants, including inherited and *de novo* variants, highly contribute to individual risk for ASD, and their analysis, considerably improved by the NGS technologies, has led to the identification of more than 100 ASD candidate genes.

This PhD project aimed to explore the contribution of rare variants in ASD susceptibility through NGS approaches in a clinically well-defined cohort of 106 ASD families including 125 ASD individuals.

In the first part of the project, I explored the contribution of inherited rare variants towards the manifestation of the ASD phenotype in a girl with a maternally inherited pathogenic *NRXNI* deletion. Indeed, exonic *NRXNI* deletions represent the prototype of ASD-associated variants whose clinical interpretation is complicated by incomplete penetrance and variable expressivity. The trio family was analysed by whole exome sequencing: this identified an increased burden of deleterious variants in the proband that could modulate the CNV penetrance and determine the disease development.

The focus of the second part of the project was to investigate the role of rare variants emerging from whole genome sequencing, both *de novo* and inherited, in ASD aetiology. To properly manage and analyse sequencing data, a robust and efficient variant filtering and prioritization pipeline was developed, and by its application a stringent set of rare recessive-acting and ultra-rare variants was obtained. As a first follow-up of this variant set, I performed a preliminary analysis focusing on *de novo* variants. This led to the identification of the most likely deleterious variants and highlighted candidate genes for further analyses.

Finally, in the third part of the project, considering the well-established involvement of calcium signalling in the molecular bases of ASD, I investigated the role of rare variants in voltage-gated calcium channels (VGCCs) genes. VGCCs mainly regulate intracellular calcium concentration, and alteration in their functionality has been correlated with enhanced ASD risk. Specifically, I functionally tested the effect of rare damaging variants identified in

CACNA1H. These data suggested that *CACNA1H* variation may be involved in ASD development by additively combining with other high risk variants.

In conclusion, this project highlights the challenges in the analysis and interpretation of variants emerging from whole genome and whole exome sequencing in ASD. A reliable and efficient filtering and prioritization pipeline facilitated the analysis step, allowing the identification of rare potentially deleterious variants. However, variant interpretation remains complicated, as the complex genetic architecture of ASD makes it difficult to define the meaning of specific gene variants. For this purpose, our data underline the importance of a comprehensive assessment of the genomic landscape of ASD individuals, even in presence of previously known pathogenic variants, moving from a genetic to a genomic perspective, shifting from a single variant analysis to an integrated view of many variants of different types, inheritance pattern, frequency and effect sizes, considering the role of protein-protein interactions.

Table of contents

1. Autism Spectrum Disorder	1
1.1 Genetic architecture	2
1.1.1 The contribution of rare sequence variants.....	4
1.2 Cellular pathways dysregulated in ASD.....	6
2. Aims of the study.....	8
3. Materials and Methods	10
3.1 Study sample	10
3.2 Whole exome and whole genome sequencing.....	10
3.3 NGS data analysis.....	12
3.3.1 Analysis of WES data from a trio with a <i>NRXNI</i> microdeletion ASD carrier	12
3.3.2 WGS data analysis.....	13
3.4 Validation of <i>NRXNI</i> microdeletion by quantitative PCR	14
3.5 Generation of WT and mutant p3xFLAG-CMV-10-CACNA1H plasmids	16
3.6 Sanger sequencing	22
3.7 Transient mammalian cells transfection with PEI.....	26
3.8 Immunofluorescence assay.....	27
3.9 Whole-cell patch-clamp.....	28
4. Results – Part I: WES analysis in a trio with a <i>NRXNI</i> microdeletion	30
5. Results – Part II: Pipeline development for NGS data filtering and prioritization and <i>de novo</i> variants analysis.....	36
5.1 <i>De novo</i> variants analysis	40
6. Results – Part III: <i>CACNA1H</i> biallelic variants in two unrelated ASD families.....	43
7. Discussion.....	52
7.1 An increased burden of rare exonic variants in <i>NRXNI</i> microdeletion carriers is likely to enhance the penetrance for autism spectrum disorder.....	52
7.2 <i>De novo</i> variants analysis	56
7.3 The contribution of <i>CACNA1H</i> biallelic variants in ASD susceptibility	59

8. Conclusions	63
Bibliography	65
Appendix	79

1. Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental disorder with an early onset, typically within the third year of life, and affecting more than 1% of the population (Maenner et al. 2020). The core symptoms are characterised by impairments in social interactions, communication difficulties, and repetitive behaviours and restricted interests. However, as the spectrum concept suggests, ASD phenotype is extremely heterogeneous due to the great variability of symptoms severity between individuals and the presence of comorbidities. Indeed, ASD usually coexists with other psychiatric, neurological and medical conditions, including intellectual disability (ID), epilepsy, language delay, attention-deficit hyperactivity disorder (ADHD), anxiety, sleep disorders and gastrointestinal problems (Bourgeron 2015, C Yuen et al. 2017).

Idiopathic ASD is the most prevalent form of the disease, while ~10% of cases presents syndromic ASD, in which ASD traits are part of more complex genetic syndrome also including congenital anomalies or dysmorphic features (Devlin and Scherer 2012). Syndromic ASD arises within both syndromes caused by some submicroscopic chromosomal rearrangements affecting clinically relevant genomic loci, and monogenic diseases, caused by single gene mutations such as Fragile X syndrome and Rett syndrome.

Advanced parental age and preterm birth have been established as risk factors for ASD (Lyll et al. 2017). Moreover, since the first studies exploring the disorder epidemiology, it has emerged that ASD prevalence in boys is greater than in girls, with a male to female diagnosis ratio of 4:1 (Maenner et al. 2020). Both hormonal and neurological factors have been suggested as causes for gender difference in ASD prevalence, leading to different hypotheses and theories. The most supported theory is the presence of a sex-related protective effect, due to which females would require a higher genetic burden compared to males to reach the diagnosis threshold, thus presenting a lower risk of developing the disease. As a direct implication, relatives of affected female would have greater risk to develop ASD or more autistic features in comparison to family members of affected males (Ferri, Abel and Brodtkin 2018).

Currently, no biomarkers are available for diagnostic or therapeutic purpose and ASD diagnosis is based on the evaluation of developmental history, behaviour and cognitive functions, according to the specific criteria of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders - DMS-5 (American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: DSM-5. 5thed. DSM-V;

2013. doi:10.1016/j.rasd.2014.07.021). Several standardized diagnostic instruments, consisting of professional observations and interviews, are commonly used to define behaviours in a consistent manner. Among them, the Autistic Diagnostic Observation Schedule (ADOS) and the Autism Diagnostic Interview-Revised (ADI-R) are considered the “Gold Standard” (Wiggins et al. 2015).

After ASD diagnosis, genetic testing is usually performed, in order to verify the presence of syndromic ASD. Testing for mutations in both *FMRI* and *MECP2* are recommended, since these genes are associated with Fragile X syndrome and Rett syndrome respectively, that are the monogenic disorders with higher incidence among ASD cases (Griesi-Oliveira and Sertié 2017). Array-CGH is also typically performed, considering that it is the established first-tier genetic test in the assessment of ASD forms caused by submicroscopic chromosomal structural variants (Battaglia et al. 2013). If no pathological variant is identified by array-CGH, whole-exome or whole-genome sequencing may be considered to further assess the involvement of sequence variants. Indeed, these methods have not yet been established as first level diagnostic tests, due to the difficulty of interpreting the results from a diagnostic point of view. However, they are likely to become so in the near future.

1.1 Genetic architecture

ASD is defined as a complex or multifactorial disease, with a strong genetic basis and possible interactions of genetic and environmental factors. Twin studies proved the crucial role of genetic factors in the aetiology of the disease. Based on the concordance rate of ASD in monozygotic and dizygotic twins and on population-based epidemiological studies, the estimated heritability of ASD, defined as the proportion of the phenotype variance due to genetic factors, is ~ 60-90% (Tick et al. 2016, Bai et al. 2019). However, despite high heritability, the genetic architecture of ASD is only partially understood and a genetic cause can be clearly identified in only 15-30% of ASD cases, principally due to ASD extremely complex and heterogeneous genetic architecture (Bourgeron 2015, Sullivan and Geschwind 2019).

Many key players are involved in the genetic heterogeneity of ASD, including common variants of small effect and rare variants of large effect, both *de novo* and inherited.

In each individual the genetic contribution to ASD can range from a single highly penetrant mutation appearing *de novo* to a combination of thousands of common low-risk variants inherited from both unaffected parents. Between these two extreme inheritance models, there is a wide range of intermediate situations, representing the majority of ASD cases, in which

ASD is the result of interactions between rare and common variants. Specifically, common and rare variants effects additively combine to determine ASD liability, even though how exactly they interplay to jointly confer risk for ASD is still under exploration (Bourgeron 2015, Klei et al. 2021).

Common variants, defined as single nucleotide polymorphisms with minor allele frequency (MAF) $\geq 1\%$, individually have just low impact on ASD phenotype, but collectively play the dominant role in the genetic liability of the disease at the population level. Common variants identification derives from genome-wide association studies (GWAS). Early GWAS in ASD proved to be underpowered and significant association for individual variants has only recently been identified, thanks to large-scale international collaborations (Grove et al. 2019).

In contrast to GWAS approaches, studies focussing on rare variants (MAF $< 1\%$) have significantly advanced our knowledge of ASD genetics. Even if it has been estimated that rare variants account for 10-25% of ASD cases (Bourgeron 2015), they contributed to the identification of more than 100 risk genes.

Rare variants include both inherited variants and variants occurring *de novo* in the proband. It has been found that *de novo* SNVs occur more frequently in the paternal germline and that the mutation rate increases with paternal age. *De novo* mutations analysis has provided persuasive evidence for several ASD risk genes (O'Roak et al. 2012, Kong et al. 2012, Krumm et al. 2014, Belyeu et al. 2021).

Moreover, several types of rare variants are implicated in ASD aetiology, including chromosomal rearrangements involving large genomic portion, copy number variants and sequence variants.

Chromosomal aberrations consist of changes in chromosome structure (structural alterations) or number (aneuploidies), detected by classical karyotyping technique. They account for ~2-5% of ASD cases and are typically identified in syndromic ASD cases with dysmorphic features or congenital anomalies. Chromosome structural alterations associated with ASD include deletions, duplications, translocations and inversions, and have been reported for every chromosome. The most frequent chromosomal abnormality is the 15q11-q13 maternal duplication, identified in 1-3% of ASD cases and already associated with Prader-Willi syndrome/Angelman syndrome. However, also several aneuploidies have been associate with ASD, such as trisomy 21 (Down syndrome), monosomy X (Turner syndrome), and aneuploidies 47 XYY and 47 XXY (Devlin and Scherer 2012).

Copy number variations, or CNVs, are defined as submicroscopic alterations of chromosomal structure, consisting in deletion or duplication ranging from kilobases to megabases in size, and decreasing or increasing the DNA content. CNVs account for the largest component of the human genetic variation and have been implicated in several diseases, including ASD (Zarrei et al. 2015). Particularly, both *de novo* and inherited CNVs have been identified in ~10% of ASD cases, with different penetrance and expressivity. *De novo* CNVs, more prevalent in sporadic ASD, typically show higher penetrance and a strong impact on the phenotype, causing more severe symptoms and sometimes playing a causative role, especially in families with a single affected individuals (simplex families) (Lai, Lombardo and Baron-Cohen 2014). In other cases, both inherited and *de novo* CNVs show incomplete penetrance and variable expressivity, having a moderate or mild effect on the phenotype. In these cases, the onset of the disease may be attributed to a multiple-hit or oligogenic model (Girirajan et al. 2012), with the occurrence of multiple rare risk variants and/or common variants polygenic risk. Indeed, some of these CNVs have also been identified in unaffected relatives, in subjects affected by other disorders and in the control population. CNVs most commonly associated with ASD involve the 15q11q13 Prader-Willi syndrome/Angelman syndrome, the 2q37 deletion syndrome, the 16p11.2 and the 22q11.2 deletion syndrome regions (Vorstman et al. 2017, Bacchelli et al. 2020, Lu et al. 2020, Mahjani et al. 2021).

1.1.1 The contribution of rare sequence variants

Sequence variants consist of small changes in DNA sequence, including both single nucleotide variants, or SNVs, and small insertion or deletion, or InDels. The advent of next generation sequencing (NGS) technologies has led to a surge in the analysis of the contribution of rare variants in ASD susceptibility. Indeed, by massively parallel sequencing of exome or genome, in whole-exome sequencing (WES) or whole-genome sequencing (WGS) approaches, NGS allows to explore thoroughly the genetic variability of the coding region or of the entire genome of large cohorts of samples.

Today, it is known that rare sequence variants, both inherited and *de novo*, highly contribute to individual risk for ASD, and high-throughput sequencing is playing an important role in the genetic characterization of the disease, supporting the idea to use it as first-tier clinical approach to investigate the involvement of both rare sequence and structural variants in ASD (Mahjani et al. 2021).

Since 2011, numerous WES studies on increasingly large cohorts of individuals have been performed, and over 100 different genes have been clearly implicated in ASD risk (Iossifov et al. 2012, Sanders et al. 2012, Neale et al. 2012, O'Roak et al. 2012, Lim et al. 2013, De

Rubeis et al. 2014, Iossifov et al. 2014, Satterstrom et al. 2020, Cameli et al. 2021, Mahjani et al. 2021). The largest WES study to date, performed in 11,986 ASD patients, led to the identification of 102 ASD genes, of which 30 are novel genes, not previously implicated in neurodevelopmental disorders (Satterstrom et al. 2020). Since the majority of disease-causing variants occur in the coding region of the human genome, initially researchers focused on the analysis of coding variants and WES studies were preferred. More recently, whole genome sequencing (WGS) studies are increasingly performed, allowing a more comprehensive assessment of rare variation in ASD with increased sensitivity over WES and the exploration of both coding and non-coding portion of the genome (Turner et al. 2016, C Yuen et al. 2017, Ruzzo et al. 2019, Wilfert et al. 2021, An et al. 2018). It is plausible that regulatory and intronic variants could contribute to disease susceptibility by altering the expression levels of ASD-associated genes or causing aberrant splicing, respectively. However, the impact of many non-coding causative variants is difficult to interpret, given the unavailability of established methods for their systematic annotation, and their most likely smaller effect compared to missense variants in the coding sequence (Darnell 2020).

Rare potentially damaging sequence variants have been found to be frequent in ASD patients, statistically correlating with more severe disease forms, in which ASD is associated with ID, epilepsy, congenital anomalies and scholastic skill disorders (Mahjani et al. 2021).

Among rare sequence variants, sporadic and high-impact *de novo* variants have been primarily and extensively studied and they still are under exploration. However, despite their key role in the genetic of ASD, *de novo* variants account for only a small portion of cases, sometimes clearly explaining the disease for simplex cases, but being less likely to explain multiplex cases (Iossifov et al. 2014, Krumm et al. 2015, Turner et al. 2017). Indeed, an enrichment of sporadic or *de novo* genetic events has been observed in simplex families, while families with multiple ASD-affected children (multiplex) are more likely to be explained by inherited disease-predisposing variants. For this reason, also the contribution of inherited variants is under investigation, particularly focusing on specific classes of sequence variants, such as likely gene-disrupting (LGD) or protein-truncating variants and missense variants predicted to be deleterious according to measures of functional severity (Satterstrom et al. 2020, Wilfert et al. 2021). Inherited risk variants, showing lower average effect size and reduced penetrance compared to *de novo* variants, are more likely to act additively in a multi-hit model. Whole-genome sequencing of multiplex ASD families allows the identification of rare inherited variants associated with disease. Interestingly, recent whole-genome sequencing studies of multiplex ASD families reported that inherited and *de novo* variants seem to

converge on related pathways, even though probably acting by different genetic mechanisms and modulating distinct sets of genes (Ruzzo et al. 2019, Wilfert et al. 2021).

Despite numerous advances in the definition of the genetic architecture of ASD and the role of the different genetic factors involved, much remains to be clarified. In addition to further defining the role of common variants, it remains to understand the mechanism by which these variants additively interact with rare variants in cumulative ASD susceptibility, and to establish whether, among rare variants, inherited and *de novo* variants actually act with distinct genetic mechanisms.

1.2 Cellular pathways dysregulated in ASD

ASD is a polygenic disorder, associated with numerous genetic risk factors. A comprehensive list of >1,000 ASD susceptibility genes, classified according to the strength of the evidence in favour of each gene, is provided by the SFARI Gene database (<https://gene.sfari.org/>). However, it is widely accepted that many ASD-related genes are still to be discovered.

Although ASD risk genes are widely expressed and are implicated in many and various biological activities, enrichment and protein-protein interaction analyses have revealed ASD risk genes converge into a small set of specific molecular pathways, including neuronal signaling and development, chromatin remodeling and transcriptional regulation, synapses formation and function, ion transport, MAPK and calcium signaling (De Rubeis et al. 2014, Wen, Alshikho and Herbert 2016, Iakoucheva, Muotri and Sebat 2019, Masini et al. 2020, Reilly et al. 2020). Additionally, more recent enrichment analysis of inherited ASD risk variants have also implicated pathways related to cell cycle and the cytoskeletal organization (Ruzzo et al. 2019).

Many ASD risk genes are key players in synaptic plasticity, modulating synaptic formation and connectivity at different levels. Mutations in these genes could give rise to alteration of the typical neuronal connectivity, thus altering the coordination and competition of neuronal networks and increasing the risk for ASD. More generally, it can be hypothesized that several ASD risk variants may result in insufficient synaptic homeostasis, a phenomenon that maintains postsynaptic stability by regulating synaptic strength, thus leading to neural activity that is too high or too low (Bourgeron 2015). ASD-associated mutations have been found in genes encoding for synaptic receptors, cell adhesion molecules, scaffold proteins, proteins involved in the formation and morphology of dendritic spine and axons, proteins engaged in the transfer of synaptic molecules to the cell surface, cycling of receptors and vesicular cargo

transport along microtubules. In addition, several genes associated with ASD are chromatin modifiers, transcription regulators or regulators under neuronal activity of protein synthesis and degradation, influencing synaptic plasticity acting on gene and protein expression, such as the transcription repressor *MECP2*, the transcription factors *MEFC2*, *CTNNB1* and *CHD8*, and genes coding for the mTOR, FMRP and ubiquitin-proteasome system pathways. A comprehensive review of main ASD candidate genes involved in synaptic architecture and functionality has been provided by Masini and colleagues (Bourgeron 2015, Masini et al. 2020, Satterstrom et al. 2020, Wilfert et al. 2021).

Among synaptic proteins mediating cell adhesion, neuroligins are essential for synaptic function, being directly involved in the maintenance of synaptic homeostasis in conjunction with neuroligins. Rare mutations in *NRXN* genes, both CNVs and SNVs, have been reported in individuals with ASD and other neuropsychiatric conditions, and *NRXN1* has been indicated among the commonest rare genetic factors implicated in ASD susceptibility (Pinto et al. 2010, Pinto et al. 2014).

Ion channels also contribute to neuronal connectivity by mediating ions transfer across the cell membrane, and thus being implicated in the action potentials generation, gene expression and neuronal cell morphology. Several genes coding voltage-gated ion channels have been implicated in ASD, including genes for voltage-gated sodium channels, potassium channels and calcium channels (Schmunk and Gargus 2013). Voltage-gated calcium channels, also indicated as VGCCs or Ca_v , regulate intracellular calcium concentration and are strictly interconnect with synaptic plasticity and synaptic homeostasis, by participating in the modulation of neurotransmitter release and modelling of synaptic membrane composition. Both common and rare variants in VGCCs have been reported in ASD individuals, convincingly indicating most of them as ASD candidate genes (Schmunk and Gargus 2013, Wen et al. 2016, Reilly et al. 2020).

2. Aims of the study

Autism Spectrum Disorder (ASD) is a multifactorial and highly heritable neurodevelopmental disorder. The genetic architecture is complex and heterogeneous, consisting of a combination of common low-risk and more penetrant rare variants. At one extreme of the genetic architecture of ASD there is polygenic risk, which is given by the sum of thousands of common genetic variants of small effect. At the other extreme are rare variants, which play a predominant role in the single individual susceptibility, with a variable effect depending on whether they are *de novo* or inherited variants. *De novo* variants have been estimated to account for up to ~30% of cases (Iossifov et al. 2014), while inherited rare coding variants with incomplete penetrance and variable expressivity have been found to explain another small portion of cases (Iakoucheva et al. 2019). Taken together, both *de novo* and inherited rare variants highly contribute to individual risk for ASD, and they offer distinctive opportunities to increase our knowledge on ASD-related biology and aetiology. Recently, rare variants analysis has considerably expanded mainly due to the application of high-throughput NGS technologies (Mahjani et al. 2021). Currently, the investigation of rare genetic variation has led to the identification of more than 100 ASD candidate genes, showing functionally convergence on a small set of common pathways (Iakoucheva et al. 2019, Satterstrom et al. 2020).

This PhD project explores the genetic landscape of ASD by investigating the role of rare variants in the genetic aetiology of ASD using NGS data. Specifically, a cohort of 106 families with at least one individual with ASD diagnosis was collected in collaboration with the UOSI Disturbi dello Spettro Autistico, IRCCS Istituto delle Scienze Neurologiche – Bologna (Italy) coordinated by Dr. Paola Visconti and genetically characterized by WGS or WES.

In the first part of this project, we investigated the genomic background in a girl with ASD from our cohort, given the presence of an inherited pathogenic microdeletion involving the *NRXNI* gene, with the goal of investigating the burden of rare variants in this proband and their contribution towards the manifestation of the ASD phenotype in *NRXNI* deletion carriers. Although *NRXNI* is one of the most well-studied ASD genes and copy number variations (CNVs) in this gene are among the commonest rare variants associated with ASD, clinical interpretation of these variants is often complicated by incomplete penetrance, variable expressivity and different phenotypic manifestations (Cosemans et al. 2020, Darnell 2020).

The second part of the project was focused on the analysis of rare variants emerging from WGS of the remaining 105 ASD families of the cohort, including 124 ASD individuals. The first goal of this part was to define a semi-automated, reliable and efficient pipeline for NGS variant filtering and prioritization. Indeed, given the large amount of variants identified by WGS, a prioritization strategy was necessary. Moreover, NGS analysis was restricted to variants occurring in the coding sequence or affecting canonical splice sites, as they are easier to interpret and more likely to be causative compared to non-coding variants. The developed pipeline was applied for the analysis of all families, leading to the identification of a set of rare and ultra-rare variants. Within ultra-rare variants, a preliminary analysis of *de novo* variants was performed to prioritize *de novo* variants in genes already implicated in ASD, or genes belonging to pathways dysregulated in ASD, to further understand their contribution to ASD susceptibility.

Then, in the third part of this work, considering the well-established involvement of calcium signalling in the molecular bases of ASD (Wen et al. 2016, Reilly et al. 2020), we investigated the role of voltage-gated calcium channels (VGCCs or Ca_v channels) in ASD susceptibility, by assessing the presence of rare coding and splicing variants in VGCCs genes in our WGS data set. VGCCs play a major role in the regulation of the intracellular calcium concentration. They stand out as a shared risk factor for five major psychiatric disorders (Consortium 2013), and alteration in VGCCs functionality has been correlated with perturbation of intracellular calcium homeostasis and enhanced ASD susceptibility (Schmunk and Gargus 2013, Liao and Li 2020, Reilly et al. 2020). Our analysis was focused on the low-voltage-activated T-type Ca_v3.2 channel gene, *CACNA1H*, in which we identified rare biallelic mutations in two unrelated families from the WGS analysis. *CACNA1H* inherited heterozygous variants have been previously identified in ASD individuals and implicated in the modulation of ASD phenotype (Splawski et al. 2006). Thus, the third goal of the project was to functionally characterize the identified *CACNA1H* mutations and clarify their involvement in the disease development in the two families.

3. Materials and Methods

3.1 Study sample

The study sample consisted of 106 unrelated families with one or more subjects affected by Autism Spectrum Disorder (ASD), including 85 simplex (one ASD individual) and 21 multiplex (two or more ASD individuals) families.

Families were recruited in collaboration with UOSI Disturbi dello Spettro Autistico, IRCCS Istituto delle Scienze Neurologiche - Bologna, Italy. ASD was ascertained using a set of standardized diagnostic instruments, evaluating the presence and severity of ASD (ADOS, CARS and M-CHAT) and assessing developmental/cognitive levels (PEP-3, Leiter-R, Griffith Scales, or Wechsler Scales) and adaptive behaviour (Vineland Adaptive Behavior Scale, VABS). Also, clinical signs like mimicry, hyperactivity, sensory abnormalities and symptoms onset were evaluated. Moreover, ASD subjects underwent array-CGH, EEG and MRI. Subclinical features in relatives were assessed using the Social and Communication Disorders Checklist and the Broad Autism Phenotype Questionnaire.

Next generation sequencing (whole genome or whole exome sequencing) analysis was performed on the entire cohort, including 125 ASD individuals (98 males and 27 females), 212 parents and 58 unaffected siblings. More specifically, whole genome sequencing was performed for 105 families, while whole exome sequencing was carried out for 1 family.

Each individual received an identification code (e.g. 2.3) consisting of family number followed by number 1, 2, 3, 4, and so on, indicating the individual was the father, mother, proband and affected or unaffected siblings respectively.

DNA samples were obtained by DNA extraction from whole blood samples using QIAmp DNA Blood Midi and Maxi kit (Qiagen).

3.2 Whole exome and whole genome sequencing

In order to perform library preparation for next generation sequencing, DNA samples were quantified by Qubit fluorometer, using Qubit dsDNA BR (Broad-Range) Assay kit (Life Technologies). DNA library preparation protocols require accurate quantification of double-strand DNA (dsDNA), in order to avoid suboptimal results or completely failure of protocol due to inaccurate initial DNA concentration measurement. Fluorometric DNA quantification is based on the use of an intercalating fluorescent dye that binds specifically dsDNA, ensuring a highly selective dsDNA measurement compared to common spectrophotometric methods (e.g. NanoDrop). Indeed, these methods generally overestimate DNA concentration and do

not allow to easily quantitate dsDNA alone, given all nucleic acid (ssDNA, RNA) absorb UV light at or near 260nm wavelengths.

Quantified DNA were used for library preparation for whole genome or whole exome sequencing. For whole genome sequencing (WGS), library preparation and sequencing were carried out at New York Genome Center (NYGC), New York, using 1µg of DNA, an Illumina PCR-free library protocol, and the Illumina HiSeq X Ten sequencing platform.

For whole exome sequencing (WES), DNA library generation and exome capture were performed in our laboratory.

Genomic DNA (gDNA) samples were diluted in elution buffer (Qiagen) in order to have 1 µg of DNA in 100 µl as total volume, and mechanically sheared by sonication to obtain fragments which average size was 180-220 bp, as required by library generation protocol. DNA fragments size was verified by microfluidics-based automated electrophoresis using Agilent 2100 Bioanalyzer (Agilent Technologies) and Agilent High Sensitivity DNA Kit (Agilent Technologies), which allows sizing just 1 µl of dsDNA sized from 50-7000 bp.

KAPA Hyper Prep Kit (Roche) and SeqCap EZ MedExome Enrichment Kit (Roche) were used for gDNA library preparation and exome capture respectively, following SeqCap EZ HyperCap Workflow (NimbleGen Roche) protocol. Size of pre-capture libraries fragments was checked by Agilent DNA 7500 Kit (Agilent Technologies).

Before sequencing, concentration and quality of the amplified captured multiplex DNA sample were determined using NanoDrop spectrophotometer. Specifically, sample should contain 500 ng of DNA, and A_{260}/A_{280} value should be 1.7-2.0. Moreover, the sample average fragment length should be between 150-500 bp. Thus, we verified sample fragment length using Agilent High Sensitivity DNA Kit (Agilent Technologies).

Sample concentration required for sequencing procedure was 1.1-1.9 nM, consequently the amplified captured multiplex DNA sample underwent to serial dilutions to reach the appropriate concentration, using Qubit dsDNA BR (Broad-Range) and High Sensitivity (HS) Assay kit (Life Technologies) for intermediate quantifications.

Pair-end reads sequencing was performed on the Illumina NextSeq550 system (Illumina Inc, San Diego, CA, USA) at the Interdepartmental Cancer Research Center "Giorgio Prodi", University of Bologna.

3.3 NGS data analysis

3.3.1 Analysis of WES data from a trio with a *NRXN1* microdeletion ASD carrier

WES for the family including a female ASD child carrying a maternal *NRXN1* microdeletion, was carried out as illustrated in section 3.2. After sequencing, coverage analysis was conducted using QualiMap (Okonechnikov, Conesa and García-Alcalde 2016), for both coverage statistics and comparison of coverage between family samples. The mean coverage depth and the mean quality mapping were 120-122X and 58 respectively for the three samples.

Data analysis was performed using the Consensus Variant Calling System (CoVaCS) workflow, hosted at the Italian computing centre Cineca (Chiara et al. 2018). CoVaCS is a multistep automated pipeline based on the integration of multiple bioinformatic tools and algorithms for genotyping, variant calling and annotation of NGS data. One of the key strengths of this pipeline is the use of three different variant callers (GATK, VarScan and FreeBayes) for variants detection (both Single Nucleotide Variants – SNVs – and small insertions and deletions – InDels) to perform a consensus variant calling and to offer a strong advantage in variants identification confidence. Indeed, results from the variant calling algorithms are combined, keeping only variants called by two out of three algorithms.

Briefly, through CoVaCS, WES raw data in FASTQ format (two FASTQ files for each sample, one for each strand) underwent trimming, quality checks and reads mapping. Reads trimming was performed using Trimmomatic, and quality checks were carried out using the Fastqc program. BWA was employed for reads alignment to reference genome hg19 (GRCh37). Alignment results were stored in BAM file that could be analysed by Integrative Genomics Viewer (IGV) software (Broad Institute). After reads alignment to the reference sequence, PCR duplicates were removed using the Picard tool suite, and joint sample variant calling was carried out, generating the consensus variants callset for SNVs and InDels converging in two VCF files respectively. Identified variants were annotated using ANNOVAR software (Wang, Li and Hakonarson 2010), adopting RefSeq for gene-based annotation.

Annotated variants were filtered and prioritized. Rare variants were selected by filtering variants according to MAF in Genome Aggregation Database (gnomAD) v.2.1.1 (<https://gnomad.broadinstitute.org/>) (Karczewski et al. 2020) and 1000 Genomes Project databases, adopting a threshold of $MAF \leq 0.5\%$ in gnomAD exome and $< 1\%$ in gnomAD genome and the 1000 Genomes Project (Auton et al. 2015).

Exonic and splicing variants were selected, and synonymous variants were excluded. Likely deleterious variants were prioritized to capture likely gene-disrupting or LGD (stop-gain, stop-loss, frameshift and canonical splice-site variants) and damaging missense variants. Six functional impact predictors were used for the definition of damaging missense variants, considering as damaging which variants meeting at least two of the following criteria: SIFT ≤ 0.05 , Polyphen2 (HDIV) ≥ 0.95 , Mutation Assessor score ≥ 2 , CADD ≥ 15 , placental mammal PhyloP ≥ 2.4 and vertebrate PhyloP ≥ 4 (Yuen et al. 2015). Finally, to remove low-quality variants, variants were required to have depth of coverage (DP) ≥ 10 , and genotype quality (GQ) ≥ 20 .

Mutation intolerant genes were defined according to RVIS (Petrovski et al. 2013) and pLI (Lek et al. 2016) scores. Genes previously associated with ASD were selected using SFARI Gene database (<https://gene.sfari.org/>, Release: 2019 Q4), while brain expressed, synaptic and postsynaptic density (PSD) genes were defined as previously described (Kang et al. 2011) (Jansen et al. 2017) (Bayés et al. 2011).

3.3.2 WGS data analysis

Raw data from WGS underwent a first-level analysis at the NYGC, where quality controls, alignment and variant calling were carried out according to the pipeline developed by the Center for Common Disease Genomics project (<https://github.com/CCDG/Pipeline-Standardization/blob/master/PipelineStandard.md>). Briefly, sequencing reads were aligned to the reference genome using BWA, and duplicate reads were marked using Picard. SNVs and InDels were called using the GATK HaplotypeCaller package version 3.5, and variant call accuracy was estimated using the GATK Variant Quality Score Recalibration (VQSR) approach. Particularly, variant calling was conducted considering samples belonging to the same family simultaneously, and calling variants jointly in the proband and in relatives. In this way, a Variant Call Format (VCF) file was obtained for each family.

Detected variants were functionally annotated in order to associate to each variant information regarding its frequency in specific databases, its consequence on the affected gene and the corresponding protein, and information about the region in which variant is included. Moreover, additional information about involved genes can be added.

WGS VCF files were annotated employing ANNOVAR software (Wang et al. 2010), using RefSeq for gene-based annotation and keeping as reference genome assembly the most recent one, hg38 (GRCh38). Both gene-based, region-based and filter-based annotation were performed, utilizing the following databases:

- refGene, knownGene and ensGene databases for the gene-based annotation;

- genomicSuperDups (for the identification of variants mapping to segmental duplications, and consequently being most likely sequence alignment errors) and cytoBand (for cytogenetic bands data) databases for the region-based annotation;
- gnomAD exome and genome v.2.1.1, gnomAD genome v.3.0, 1000g2015aug, and exac03nonpsych databases for allele frequencies annotation from public databases;
- avsnp150 (from dbSNP150) database for variants annotation;
- dbnsfp41a and revel (for REVEL score) databases for annotation about variants functional prediction;
- intervar_20180118 (from InterVar) and clinvar_20200316 (from Clinvar) databases for clinical interpretation of variants;
- dbcsnv11 and spliceai_filtered for splice site prediction data;

Moreover, a customised version of the cross-reference gene_fullxref file was employed to annotate additional information about genes, such as previously implication in ASD through SFARI Gene database (<https://gene.sfari.org/>) and in other Neurodevelopmental Disorders (NDDs), intolerance to functional mutations.

Annotated variants filtering and prioritization were performed using a customised pipeline described in section 5.

3.4 Validation of *NRXN1* microdeletion by quantitative PCR

The presence of *NRXN1* microdeletion emerged from array-CGH analysis and its segregation in family 2 was assessed by real time or quantitative PCR (qPCR) reaction, using SsoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad).

SYBR Green is a non-specific fluorescent dye binding the minor groove of double-strand DNA (dsDNA), thus allowing to track DNA amplification in qPCR experiment. Briefly, DNA quantification occurs at the end of the elongation step of each PCR round, when SYBR Green dye binds dsDNA, resulting in an increase in the level of fluorescence that is proportional to the quantity of dsDNA produced. The analysis of qPCR data is based on the threshold cycle or C_t value, defined as the cycle number at which the fluorescence rises significantly above the background fluorescence, crossing the fluorescence threshold.

qPCR primers were designed from genomic DNA sequence using Primer3 software (<https://bioinfo.ut.ee/primer3-0.4.0/>) and setting the following conditions, in addition to the common recommendations for optimal PCR primers design:

- Product size range = 90-140 bp;

- Primer size = Min: 20, Opt: 22, Max: 30;
- Primer GC% = Min: 40, Max:50;
- Max self-complementarity = 5;
- Max 3' self-complementarity = 2;
- Max Poly-X = 3.

For the *NRXN1* microdeletion validation we used five pairs of primers, listed in Table 1. Two primers pairs mapped inside the deleted region, allowing the CNV validation, while the other two mapped outside the CNV, overlapping the closest not deleted array-CGH probes for CNV boundaries definition. Moreover, another pair of primers was used to amplify a control region on the *FOXP2* gene at 7q31.1.

The CNV presence was assessed in all family individuals and in a control subject predicted to have none *NRXN1* copy number alteration, setting up each reaction in triplicate and preparing reaction mix as indicated in Table 2. Moreover, a negative control was included for each reaction.

qPCR was performed using CFX Connect™ Real-Time PCR Detection System (Bio-Rad) and CFX Manager software, according to the following protocol:

- Polymerase activation: incubation at 98°C for 3 min;
- DNA denaturation: incubation at 98°C for 10 sec;
- Annealing and elongation: incubation at 60°C for 30 sec;
- Melting curve generation: incubation at 65°C for 5 sec, increasing temperature by 0.5°C every cycle, until reaching incubation at 95°C.

Specifically, melting or dissociation curve generation step consists in the fluorescent signal measurement while temperature is gradually increased. The analysis of melting curve allows to ascertain reaction specificity, detecting non-specific PCR product and/or primers dimers that could reduce PCR efficiency. Indeed, SYBR Green dye binds dsDNA in non-specific way, binding also primers dimers and non-specific PCR product possibly present in reaction mix and giving spurious fluorescent signals. By analysing melting curve, signal peaks not overlapping the melting temperature, T_m , of PCR product indicates the presence of primers dimers (T_m lower than PCR product T_m) or non-specific PCR product (T_m different from PCR product T_m).

For each sample, qPCR data were compared to data of the *FOXP2* gene, used as endogenous control gene, and to data of the control subject, used as external control.

The number of copies of each amplified fragment was calculated using the $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen 2001), as indicated by the following formula.

$$\Delta Ct_{sample\ DNA} = Ct_{amplified\ fragment} - Ct_{FOXP2}$$

$$\Delta Ct_{control\ DNA} = Ct_{amplified\ fragment} - Ct_{FOXP2}$$

$$\Delta\Delta Ct = \Delta Ct_{sample\ DNA} - \Delta Ct_{control\ DNA}$$

$$Copy\ number = 2 \times 2^{-\Delta\Delta Ct}$$

Table 1. Primers used for validation and segregation analysis of NRXN1 microdeletion.

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (bp)	Fragment position (hg19)
NRXN1_EX24_F	ttatagggtagggtgttctgc	22	111	chr2:50149639-50149749
NRXN1_EX24_R	agcattgtcccaataaactcc	22		
NRXN1_EX23_F	cactcactctgccagcaata	22	100	chr2:50170343-50170442
NRXN1_EX23_R	caacgagtctacctgtgttg	22		
NRXN1_Int6_A_F	aactcacatttctaccgcaag	22	139	chr2:50934420-50934558
NRXN1_Int6_A_R	ttgaatcagaaccagttagcc	22		
NRXN1_Int6_B_F	ccaagtgtgggtgatgattta	22	101	chr2:51069239-51069338
NRXN1_Int6_B_R	tcttctgggcacctcatct	20		
FOXP2_F	tgctagaggagtgggacaagta	22	140	chr7:114121559-114121698
FOXP2_R	gaagcaggactctaagtgcaga	22		

Table 2. qPCR reaction mix components.

Component	Quantity per reaction (Reaction volume 15 μ l)
PCR-grade water	1 μ l
SsoAdvanced™ Universal SYBR® Green Supermix 2X	7.5 μ l
Forward primer (10 μ M)	0.75 μ l
Reverse primer (10 μ M)	0.75 μ l
Template DNA (25 ng)	5 μ l

3.5 Generation of WT and mutant p3xFLAG-CMV-10-CACNA1H plasmids

Plasmidic constructs for the expression of *CACNA1H* recombinant proteins in mammalian cells (HEK-293T) were generated by cloning *CACNA1H* coding sequence (NM_021098.3) in the p3xFLAG-CMV-10 mammalian expression vector (Sigma-Aldrich), obtaining plasmid DNA encoding an N-terminal 3xFLAG fusion protein detectable by anti-FLAG antibody. *CACNA1H* coding sequence derived from a1Ha-pcDNA3 plasmid (Addgene plasmid # 45809 ; <http://n2t.net/addgene:45809> ; RRID:Addgene_45809) (Cribbs et al. 1998).

Cloning procedure is described in section 6, while employed technologies are described below.

Restriction enzyme digestion reaction

Restriction enzyme digestions were widely performed during cloning process to shuttle *CACNA1H* coding sequence into the expression vector and to assess the success of each cloning step (diagnostic restriction digest). Moreover, template for whole-plasmid PCR reaction for site-directed mutagenesis was obtained by restriction enzyme digestion cloning strategy.

Reactions were set up selecting suitable restriction enzymes (New England BioLabs) for single or double digestion, adjusting units of enzymes to the amount and purity of plasmid DNA to digest (10 enzyme units/ μg of DNA). Digestion buffer (NEBuffer) assuring major enzyme % activity was selected. Reactions were incubated at 37 °C, the optimal incubation temperature for employed enzymes, for 1-3 hours.

Results of digestion reaction were visualized and analysed by gel electrophoresis using 0.8% TBE agarose gel containing Midori Green Advance stain (Nippon Genetics) and employing 1 kb Plus DNA ladder (New England BioLabs) as size standard. Gel Doc XR+ System (Bio-Rad) was used for gel imaging.

PCR for 5'-CACNA1H fragment cloning in p3xFLAG-CMV-10-3'-CACNA1H

To clone the entire *CACNA1H* CDS in the 3xFLAG expression vector, 5'-CACNA1H fragment was amplified from a1Ha-pcDNA3 vector by PCR, and then it was sub-cloned into the previously obtained p3xFLAG-CMV-10-3'-CACNA1H. PCR was performed using Herculase II Fusion DNA Polymerase (Agilent Technologies), a highly processive polymerase allowing the amplification of targets >10 kb. Primers (reported in Table 3) were designed from the gene coding sequence in order to amplify a 4268 bp fragment, from gene start site to the cut site of EcoRV, that was used for the previous 3'-CACNA1H CDS fragment cloning. For forward primer design, a substitution ATG>C was introduced at 5'-gene start site to keep the gene CDS in frame with 3xFLAG.

Manufactory directions for targets 1-10 kb were followed for reaction mix setting, that is described in Table 4. Amplification was conducted using a touch-down PCR program with 68 °C and 63 °C as annealing temperature for total 40 cycles and setting cycling conditions on target length and template DNA type.

PCR products were analysed by agarose gel electrophoresis, as described for the digestion reaction products.

Table 3. Primers employed for 5'-CACNA1H fragment amplification from a1Ha-pcDNA3.

Primer name	Sequence (5'-3')	Primer size (bp)	Product size (bp)
CACNA1H_upperFragm_F	caccgagggcgcacggg	17	4268 bp
CACNA1H_upperFragm_R	atcagcgtctccaccaccagc	21	

Table 4. Reaction components for 5'-CACNA1H fragment amplification from a1Ha-pcDNA3.

Component	Quantity per reaction (Reaction volume 25 μ l)
PCR-grade water	9.75 μ l
5X Herculase II reaction buffer	5 μ l
dNTPs mix (2.5 mM each dNTP)	2.5 μ l
Forward and reverse primer mix (1 μ M each primer)	6.25 μ l
Herculase II fusion DNA polymerase	0.5 μ l
Template DNA (5 ng)	1 μ l

Whole-plasmid PCR

WGS identified variants were introduced to the WT *CACNA1H* coding sequence by site-directed mutagenesis, performing whole-plasmid PCR reaction with Herculase II Fusion DNA Polymerase (Agilent Technologies, Inc.) and using divergent primers, whose sequence is reported in Table 5. Indeed, whole-plasmid PCR allows the amplification of the entire plasmid, through divergent primers going around all the plasmid and pointing away from each other, generating a linear PCR product. Moreover, it is possible to introduce simultaneously tailored mutation (insertion, deletion and single nucleotide mutation) depending on how divergent primers are designed.

In this work we wanted to introduce SNVs in template DNA, thus we designed divergent primers binding the template DNA to the mutation site, but with one of the two oligonucleotides (the forward or the reverse one) having a mismatch at the 5' end, or around it, due to the insertion of the alternative allele instead of the reference one at the mutation site. Four different whole-plasmid PCR reactions were set up (as described in Table 6), one for each mutant plasmid we wanted to generate and thus containing a different primers pair. As template DNA, two different p3xFLAG-CMV-10 types were used: one containing 5'-CACNA1H fragment for the generation of the Lys785Met and Pro849Ser mutants, and the other containing 3'-CACNA1H fragment for the generation of the Pro2124Leu and Ser2338Phe mutants.

Amplification was conducted using a touch-down PCR program with 65°C and 60°C as annealing temperature for total 30 cycles and setting cycling conditions on target length and template DNA type.

Whole-plasmid DNA amplification success was evaluated by agarose gel electrophoresis, as already described, while the presence of the desired mutations at mutation site and the absence of undesired changes along the plasmid DNA sequence were confirmed by Sanger sequencing (described in section 3.6).

Table 5. Sequence of divergent primers employed for site-directed mutagenesis through whole-plasmid PCR. Mismatch sites corresponding to changes introduced in plasmid DNA sequence through whole-plasmid PCR are reported in bold.

Primer name	Sequence (5'-3')	Primer size (bp)
CACNA1H_K785M_F	gctgcgccgcatcgtgga	18
CACNA1H_K785M_R	atg ccgctgaaggtaccagag	23
CACNA1H_P849S_F	gctctctgggctacatccggaac	23
CACNA1H_P849S_R	cgcaggccagcagctca	18
CACNA1H_P2124L_F	ggtggccggcgcgagc	17
CACNA1H_P2124L_R	agagaggcttcggggccatg	20
CACNA1H_S2338F_F	ccctcagccaccctg	17
CACNA1H_S2338F_R	aac ctggttctccagaggac	22

Table 6. Whole-plasmid PCR reaction mix composition.

Component	Quantity per reaction (Reaction volume 25 µl)
PCR-grade water	9.75 µl
5X Herculase II reaction buffer	5 µl
dNTPs mix (2.5 mM each dNTP)	2.5 µl
Forward and reverse primer mix (1 µM each primer)	6.25 µl
Herculase II fusion DNA polymerase	0.5 µl
Template DNA (5 ng)	1 µl

DNA fragments extraction from agarose gel

After electrophoresis analysis, DNA fragments resulting from digestion or amplification reactions were recovered and purified from agarose gel using QIAquick Gel Extraction kit (Qiagen), based on the employment of spin-column equipped with silica membrane for DNA absorption. DNA was eluted with elution buffer and quantified using NanoDrop.

Dephosphorylation of 5' end of digested plasmid by CIP

Before ligation reaction, 5' end of receiving plasmid DNA linearized by restriction enzyme digestion were dephosphorylated in order to prevent its re-ligation. We used calf intestinal alkaline phosphatase or CIP (New England BioLabs), that was incubated with plasmid DNA at 37 °C for 30 min. Specifically, 5 units of enzymes were used for every 1 pmol of DNA ends.

Phosphorylation of 5' end of PCR product by T4 PNK

Prior to perform ligation reaction employing PCR product (both 5'-CACNA1H fragment and plasmid DNA resulting from whole-plasmid PCR), the addition of 5'-phosphates was necessary to allow subsequent ligation. Phosphorylation reaction was conducted using T4 Polynucleotide Kinase or T4 PNK (New England BioLabs), requiring ATP as source of P_i. Reaction was set up using 10 units of T4 PNK for up to 300 pmol of 5' ends and using 1X T4 DNA Ligase Buffer (already equipped with 1 mM ATP) as reaction buffer, according to manufactory protocol. Incubation was carried out at 37 °C for 1 hour.

Purification of both dephosphorylation and phosphorylation products

QIAquick PCR purification kit (Qiagen) was employed for direct purification of both dephosphorylation and phosphorylation products. Indeed, both CIP and T4 PNK removal are necessary to promote plasmid-insert ligation efficiency and penalise plasmid re-ligation.

Ligation reaction

Ligation reaction, allowing both the introduction of the insert in plasmid and the circularization of linearized plasmid or whole-plasmid PCR products, was carried out using T4 DNA Ligase (New England BioLabs), catalysing phosphodiester bond formation between 5' phosphate and 3' hydroxyl ends. Therefore, phosphorylated 5'end and ATP-rich buffer (10X T4 DNA Ligase Buffer) are required for the reaction. We used 1 µl of T4 DNA Ligase, 50 ng of vector and a molar ratio of 1:3 vector to insert, in 20 µl reaction. Incubation was performed at 16 °C overnight. Then, 15 µl of the reaction (corresponding to ~35 ng of ligation product) were transformed into 100 µl competent cells.

Transformation of chemically competent *E. coli* cells

During cloning procedure, transformation process was conducted for plasmids propagation, using heat-shock protocol and DH5α chemically competent *E. coli* cells (cells genotype: F-, endA1, glnV44, thi-1, recA1, relA1, gyrA96, deoR, nupG, Φ80dlacZΔM15, Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ-).

Competent cells aliquot was thawed on ice for 15 min. Then, ~35 ng of the ligation product were added to the cell aliquot, mixing cells and DNA by carefully flicking the tube. The mixture was placed on ice for 30 min, and next incubated at 42 °C for exactly 30 sec for heat shock. Immediately after, the mix was placed on ice again for 5 min. Bacterial cell recovery and antibiotic resistance expression were induced by adding 900 µl of room temperature SOC to the mix and incubating it with shaking at 37 °C for 60 min. p3xFLAG-CMV-10 expression vector confers resistance to ampicillin antibiotic to transformed cells. Consequently, after cells outgrowth, 4 parts of the transformation mix were spread onto LB-agar containing ampicillin and previously warmed at 37 °C. Plate was incubated at 37 °C overnight.

The day after, a variable number (generally 3-10 colonies were picked, depending on the number of colonies on the control plate) of single colonies per plate were inoculated to generate liquid cultures of single-clone colonies of bacteria for plasmid DNA isolation. In detail, a single colony from LB-agar plate was transferred into 2 ml, 5 ml or 100 ml of liquid LB-ampicillin (liquid culture volume depends on the procedure used for the following plasmid DNA isolation), and incubated with shaking at 37 °C for 12-16 hours, given employed plasmids were high copy plasmids.

Plasmid DNA isolation: Alkaline lysis and ethanol precipitation miniprep

Miniprep via alkaline lysis and ethanol precipitation was carried out to isolate plasmid DNA for diagnostic restriction digestion along cloning process. Once positive results from diagnostic restriction digestion were obtained, new overnight cultures were generated to perform mini- or midiprep through silica membrane technology, assuring higher purified and concentrated plasmid DNA.

2 ml overnight cultures are sufficient for alkaline lysis and ethanol precipitation miniprep. 1.8 ml of the stock culture were transferred into a 1.5 ml tube and centrifuged at 4700 g for 5 min at room temperature. The supernatant was poured off, and the bacterial pellet was resuspended in 100 µl of cold Solution I (for the composition see Table 7) previously completed with RNase A $C_i=50$ µg/ml, by vortexing the solution. Then, 200 µl of Solution II (Table 7) were added and the tube content was mixed by inverting the tube carefully 7-8 times. After incubation at room temperature for 3 min, 150 µl of Solution III (Table 7) were added to the tube, that was inverted carefully 15-20 times again. The mix was incubated on ice for 5 min and centrifuged at 16200 g for 15 min at 4 °C. The supernatant was collected into a new tube by pipetting and without disturb the white precipitate containing proteins, cell fragments, salt and bacterial DNA. 700 µl of 100% ethanol were added to the supernatant for

the DNA precipitation. The mix was vortex, incubated at room temperature for 5 min and centrifuged at 16200 g for 15 min at 4 °C. The supernatant was removed and the pellet was washed with 600 µl of cold 70% ethanol. The tube was vortexed and centrifuged at 16200 g for 10 min at 4 °C. The supernatant was removed by pipetting and the pellet was air dried for ~ 15 min. Then, the pellet was resuspended with 50 µl (for high copy plasmids) of PCR-grade water and the sample was quantified by NanoDrop.

Table 7. Composition of solutions used in alkaline lysis and ethanol precipitation miniprep.

Solution I (Resuspension buffer) Store at 4 °C	Solution II (Denaturing solution) Store at room temperature	Solution III (Renaturing solution) Store at 4 °C
50 mM glucose	200 mM NaOH	5 M KOH
10 mM EDTA	1% SDS	3 M CH ₃ COOH, pH 4.8
25 mM TRIS-HCl, pH8		

Plasmid DNA isolation: Alkaline lysis miniprep and midiprep – silica membrane technology

Highly pure plasmid DNA preparations from small-scale bacterial culture were obtained using NucleoSpin Plasmid/Plasmid NoLid-kit (Macherey-Nagel), following the manufactory protocol for the isolation of high-copy plasmid DNA. In these cases, 5 ml of starter bacterial cultures were employed.

To obtain higher plasmid DNA yields, required for successive analysis, also midiprep based on silica membrane technology was performed starting from 100 ml of overnight culture and using NucleoBond Xtra plasmid purification kit (Macherey-Nagel).

3.6 Sanger sequencing

Sanger sequencing was used to both inspect CACNA1H plasmids sequence during cloning, and validate *de novo* variants emerging from NGS analysis.

PCR assay for Sanger sequencing

Sanger sequencing workflow usually includes a PCR step for the amplification of the region of interest before the sequencing reaction of high complexity DNA. For plasmids sequence inspection during cloning procedures, the amplification step was not carried out given that the sequencing target coincided with the cloned insert itself.

For *de novo* variants validation, the DNA region around the variant was amplified from gDNA. The PCR reaction was performed in all family members, including family individuals who did not undergo WGS or WES. Indeed, in these cases Sanger sequencing allowed to validate the variant and check its segregation within the family.

Primers were designed using Primer3 software (<https://bioinfo.ut.ee/primer3-0.4.0/>), following typical recommendations for optimal PCR primer design and preferring primers pairs allowing amplification of a 200-500 bp region including the variant to validate. For each pairs of primers, the optimization of amplification conditions were carried out by altering annealing temperature, MgCl₂ concentration or PCR cycle numbers.

PCR assays were set up as indicated in Table 8, using a touch-down PCR program, described in Table 9, to promote exclusive amplification of target region by improving reaction specificity. More specifically, the first ten PCR cycles were performed using an higher annealing temperature (T₁) and decreasing it by 0.5 °C every cycle to a lower annealing temperature (T₂), in order to ensure specific primer annealing; thus, the following 30 cycles were performed using T₂ as annealing temperature.

PCR products were evaluated by gel electrophoresis on 2% agarose gel containing Gel Red Nucleic Acid stain (Biotium). 100 bp DNA ladder (New England BioLabs) was used as size standard for sizing of PCR fragments. Gel Doc XR+ System (Bio-Rad) was used for gel imaging and analysis.

Table 8. Composition of PCR mix for the amplification of target region for Sanger sequencing.

Component	Quantity per reaction (Reaction volume 15 µl)
PCR-grade water	up to 15 µl
10X AmpliTaq Gold buffer	1.5 µl
MgCl ₂ (Cf = 1.5-3.0 mM)	0.9 - 1.8 µl
dNTPs mix (25 mM each dNTP)	0.12 µl
Forward primer (10 µM)	0.4 µl
Reverse primer (10 µM)	0.4 µl
AmpliTAq Gold DNA Polymerase (5 u/µl, LifeTachnologies)	0.05 µl
Template DNA (30 ng)	3 µl

Table 9. PCR program for the amplification of target region for Sanger sequencing.

Step		Temperature and time
Initial denaturation		95 °C for 15 min
Touch-Down (10 cycles: T ₁ decreases by 0.5 °C every cycle)	Denaturation	95 °C for 30 sec
	Annealing	T ₁ for 30 sec
	Extension	72 °C for 30 sec
Amplification (30 cycles)	Denaturation	95 °C for 30 sec
	Annealing	T ₂ for 30 sec
	Extension	72 °C for 30 sec

PCR clean-up

Before sequencing reaction, enzymatic PCR clean-up was performed to remove unincorporated primers and dNTPs that could interfere with sequencing. Particularly, we used Illustra ExoProStar 1-Step (Cytiva) technology, that combines Exonuclease I and Alkaline Phosphatase enzymes in a single mix for excess primers digestion and nucleotides dephosphorylation respectively. 1 μ l of enzymes mix was added to 6 μ l of PCR products, and incubated at 37 °C for 15 min (treatment step). Then, reaction mix was incubated at 80 °C for 15 min for heat inactivation of the enzymes.

Sequencing reaction

Sanger sequencing was carried out using BigDye Terminator v1.1 Cycle Sequencing kit (Applied Biosystems).

For *de novo* variants validation, sequencing reaction was performed using one out of two primers used for preceding PCR step. Forward or reverse primer was preferred according to GC content and distance from variants site. Reaction mix was prepared as indicated in Table 10, using variable PCR amplicon quantity, depending on DNA signal intensity.

For plasmids sequencing, fourteen reverse primer (listed in Table 11) were designed from plasmid sequence in order to sequence the entire insert and insert-vector junction regions. Moreover, when necessary, commonly used primer for 3xFLAG vector sequencing, i.e. CMV24 reverse primer, was used. Sequencing mix was set up as illustrated in Table 10, using 250 ng of plasmid DNA as template.

Both for *de novo* validation and plasmids sequence inspection, Sanger sequencing were performed in 96-well plate, using the following program:

Parameter	Stage/step			
	Incubate	28 cycles		
		Denature	Anneal	Extend
Temperature	96°C	96°C	50°C	60°C
Time (mm:ss)	01:00	00:10	00:05	04:00

Table 10. Sanger sequencing reaction mix setting.

Component	<i>De novo</i> variants validation	Plasmid DNA sequence inspection
	Quantity per reaction (Reaction volume 10 µl)	Quantity per reaction (Reaction volume 10 µl)
PCR-grade water	up to 10 µl	5.59 µl
BigDye Terminator Sequencing Buffer	1.75 µl	1.75 µl
BigDye Ready Reaction Mix	0.5 µl	0.5 µl
Forward or Reverse Primer (10 µM)	0.16 µl	0.16 µl
Template DNA	variable	250 ng

Table 11. Primers used for plasmid DNA sequencing.

Primer name	Sequence (5'-3')	Primer size (bp)
Seq1_CACNA1H_R	ctttctgctcacgacggg	18
Seq2_CACNA1H_R	gttccaccgagatgcagg	18
Seq3_CACNA1H_R	ggctgagactgggggag	17
Seq4_CACNA1H_R	cagcgtgtcctcatgat	18
Seq5_CACNA1H_R	aaggtgatgaagaggtcga	19
Seq6_CACNA1H_R	caatgggcctgagtgat	17
Seq7_CACNA1H_R	gcttgtagggctccagca	18
Seq8_CACNA1H_R	tgcattccaggaatggtga	18
Seq9_CACNA1H_R	atgaagagcatgagcagc	18
Seq10_CACNA1H_R	acgtctgcgtgaattca	18
Seq11_CACNA1H_R	tgcacagactctgcgtc	17
Seq12_CACNA1H_R	aggagcccacgatgatg	17
Seq13_CACNA1H_R	agacgaagaagcacagc	17
Seq14_CACNA1H_R	gaagaagaccgtggccg	17
CMV24_R	tattaggacaaggctggtgggcac	24

Sequencing reactions purification

Purification of sequencing reactions is necessary to remove salts, unincorporated dye ddNTPs and dNTPs, that could compromise the quality of the sequence and interfere with base-calling.

We purified sequencing products by ethanol precipitation, following the protocol below.

- Preparation of an appropriate volume of the “precipitation mix”, composed by 50 µl of ice-cold 100% ethanol, 10 µl of water and 2 µl of Sodium Acetate 3M pH 5.2. The precipitation of the entire volume (10 µl) of each sequencing product requires 62 µl of precipitation mix.
- The precipitation mix was added to each sequencing products in plate. Mixing by repetitive pipetting is required.
- The plate was centrifuged with seal at 2000 g for 30 min at 4 °C.

- The seal was removed and the supernatant was discarded by inverting the plate.
- Ethanol wash was performed by adding 70 µl of ice-cold 70% ethanol to each sample.
- The plate was centrifuged with seal at 1650 g for 15 min at 4 °C.
- The seal was removed and the supernatant was discarded by inverting the plate.
- The plate was centrifuged inverted at 185 g for 1 min at 4 °C.
- The plate was dried at room temperature to remove ethanol traces.

Capillary electrophoresis and Sanger sequencing analysis

Products of sequencing reaction were analysed by capillary electrophoresis, that was carried out using the ABI PRISM 3730 DNA analyser (Life Technologies). Purified samples were resuspended in 20 µl of Injection Solution (DNA Sequencing Reaction Cleanup kit, Millipore) and loaded onto the analyser.

Sequences analysis, including assemble to reference sequence and variants detection, was performed with Sequencher software (Gene Code Corporation).

3.7 Transient mammalian cells transfection with PEI

In order to perform immunofluorescence assay, transient mammalian cells transfection was carried in Human Embryonic Kidney-293T (HEK-293T) cells.

Cells were transiently transfected in 6-well plate using PolyEthylenImine (PEI), transfecting 3 µg of plasmid DNA subtype *per* well and using 0.3 µl of PEI *per* 1 µg of plasmid DNA to transfect.

Specifically, cells were plated on glass coverslips in 6-well plate and grown in Dulbecco's Modified Eagle Medium (DMEM) high glucose supplemented with 10% fetal bovine serum and 0.05 mg/ml penicillin-streptomycin, at 37 °C in a 5% CO₂ humidified atmosphere. After 24 h, a PEI mix and DNA mix (ratio of 1:1) were set up for each transfection as described in Table 12, vortexed for 30 sec and incubated at room temperature for 20 min, keeping PEI in the dark. After the incubation, 50.9 µl of PEI mix were added to each DNA mix. Obtained transfection mix were vortexed for 30 sec and incubated at room temperature for 30 min in the dark. In the meanwhile, growth medium was removed and transfection medium (or PEI medium, consisting of serum-deprived medium) was added to cells for a wash. At the end of incubation, 1.4 ml of PEI medium were added to each transfection mix, and completed transfection mix was added to cells, from which wash PEI medium was previously removed. 1.5 h after transfection, transfection mix was removed and cells were maintained in growth medium at 37 °C in a 5% CO₂ humidified atmosphere.

Table 12. Setting of transfection mix for transient PEI transfection.

	p3xFLAG-CMV-10 (empty backbone)	p3xFLAG-CMV-14-ABCC3 (positive control)	p3xFLAG-CMV-10-CACNA1H WT	p3xFLAG-CMV-10-CACNA1H Lys785Met	p3xFLAG-CMV-10-CACNA1H Pro849Ser	p3xFLAG-CMV-10-CACNA1H Pro2124Leu	p3xFLAG-CMV-10-CACNA1H Ser2338Phe	pCIG3 (empty backbone containing GFP)
PEI mix*	50 μ l							
DNA mix	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g	- 150mM sterilised NaCl 50 μ l - plasmid DNA volume for 3 μ g
PEI medium	1.4 ml							

*PEI mix	1 reaction
150mM sterilised NaCl	50 μ l
PEI	0.9 μ l (0.3 μ l per μ g of plasmid DNA to transfect)

3.8 Immunofluorescence assay

Immunofluorescence (IF) assay was carried out to assess the expression of generated CACNA1H constructs (both WT and mutants) in mammalian cells and to verify the cellular localization of the coded recombinant proteins.

HEK-293T cells were grown and transiently transfected as described in previous section. Before cells seeding, glass coverslips were washed through 100% ethanol, coated with 200 μ l of collagen ($C_f = 50$ ng/ μ l) and incubated at room temperature for 1 h, to improve cells adhesion. After incubation, a 1X Phosphate Buffered Saline (PBS) wash was conducted twice.

A GFP-coding plasmid, the pCIG3, was used as transfection control, while a p3xFLAG-CMV-10 empty vector and a 3xFLAG plasmid encoding the 3xFLAG-ABCC3 fusion protein were used as IF negative and positive control, respectively (Table 12).

48 h after transfection, cells were washed twice with 1X PBS and fixed for 15 min with 1 ml of 4% paraformaldehyde (PFA) in PBS. To remove excess of 4% PFA+1X PBS, 1X PBS wash was carried out three times. Fixed cells were blocked with 200 μ l of blocking solution, consisting of 4% normal donkey serum (NDS) and 0.05% tween in PBS, for 30 min at room temperature. Then, cells were incubated with 100 μ l of 1:200 mouse anti-FLAG M2 antibody (Sigma-Aldrich) for 1.5 h at room temperature. After three washing step with 0.05% tween-

PBS, cells were incubated with 100 μ l of 1:400 goat anti-mouse Cy3 antibody (Jackson ImmunoResearch) for 1 h at RT. Samples were washed one time with PBS and nuclei were stained with 100 μ l of Hoechst (Sigma-Aldrich, $C_f = 1 \mu\text{g/ml}$). Coverslips were mounted on glass slides through mounting medium (9-part glycerol and 1 part PBS, pH = 8.5-9.0) and image acquisitions were taken by Nikon 90i wide-field fluorescence microscope. RAW images were processed into TIF files using ImageJ open-source software.

3.9 Whole-cell patch-clamp

Electrophysiological analysis were carried out in collaboration with the Prof. Sergio Fucile at the Department of Physiology and Pharmacology, Sapienza University, Rome. HEK-293 cells were grown in DMEM supplemented with 10% heat-inactivated FBS and 1% penicillin-streptomycin, at 37 °C in a 5% CO₂ humidified atmosphere. Cells were plated on cover slides (8×10^4 cells/ml) and, after 24 h, they were transiently transfected using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocol, and adding 0.5 μ g of plasmid DNA subtype per well. Recordings were carried out 24–36 h following transfection. Electrophysiological experiments were performed using the whole-cell configuration of the patch-clamp technique. Recordings were obtained using a HEKA EPC800 amplifier, Digidata 1322A analog-to-digital converter, and pClamp 10 software (Molecular Devices, Union City, CA). Data were filtered at 2 kHz and digitized at 5 kHz. Normal external solution contained: 140 mM NaCl, 2.8 mM KCl, 2 mM MgCl₂, 2 mM CaCl₂, 10 mM HEPES, and 10 mM glucose (pH 7.4; 300 mosM). The internal pipette solution contained: 140 mM CsCl, 5 mM BAPTA, 2 mM Mg-ATP, and 10 mM HEPES (pH 7.4; 300 mOsm). Borosilicate glass pipettes were pulled with a Narishige puller to a typical pipette resistance of 3-4 M Ω . Cell capacitance was measured for each cell and access resistance compensated to 70%.

The current-voltage protocol stepped the cell membrane potential from -120 mV to test potentials starting at -110 mV and increasing to 20 mV in 10 mV increments. Test potentials were 100 ms in duration, and the membrane potential was returned to -120 mV for 10 s between acquisitions to allow complete recovery from inactivation. Peak inward Ca²⁺ currents were plotted as a function of the test potential to generate current-voltage relations (I-V). The peak currents were also normalized by the individual cell capacitance measurement for the comparison of current densities. Mean current density-V relations were fit with a modified form of the Boltzmann equation, where $I_{\text{peak}} = (V - E_{\text{rev}}) G_{\text{max}} (1/1 + \exp((V_h - V)/S))$, and E_{rev} is the reversal potential, V_h is the half-activation potential, G_{max} is the maximum slope

conductance, and S is the slope factor that is inversely proportional to the effective gating charge. To assess the voltage dependence of inactivation, the cell membrane was stepped from a holding potential of -120 mV to conditioning potentials 1 s in duration between -120 mV and -60 mV in 10 mV increments before proceeding to a test potential of -40 mV for 100 ms, from which the resulting inward Ca^{2+} currents were analysed. The voltage of half-inactivation (V_i) was estimated from Boltzmann fits of I/I_{max} versus voltage where $I/I_{\text{max}} = 1/(1 + \exp(z*(V - V_i)/25.6))$. Clampfit 10 was used to analyse all data obtained in Clampex (Molecular Devices). Current kinetics were evaluated at -40 mV test potential by measuring the time from basal to peak current, and by fitting the current decay with a single exponential equation. Fits of the I-V relations, activation and inactivation curves, and decays were carried out in SigmaPlot (Jandel Scientific). Data are presented as the means \pm S.E. Statistical tests was done with one-way analysis of variance (ANOVA).

4. Results – Part I: WES analysis in a trio with a *NRXN1* microdeletion

Array-CGH data were available for each ASD individual belonging to our collection. Array-CGH is currently the recommended first tier clinical genetic test for the identification of disease-causing CNVs in ASD or other neurodevelopmental disorder individuals.

Among ASD individuals of our cohort, the female proband of a trio family had a pathogenic deletion of ~811 kb at 2p16.3 (NC_000002.11:g.50170766_50982172del) involving exons from 7 to 23 of the *NRXN1* gene (NM_001135659.2) (Figure 1).

To confirm the presence of the CNV in the proband and assess its segregation, qPCR was performed in all family members, confirming the presence of the *NRXN1* intragenic deletion in the proband and showing that it was inherited from the unaffected mother (Figure 2). Other three rare CNVs were present in the proband, but none of them is considered to be clinically relevant.

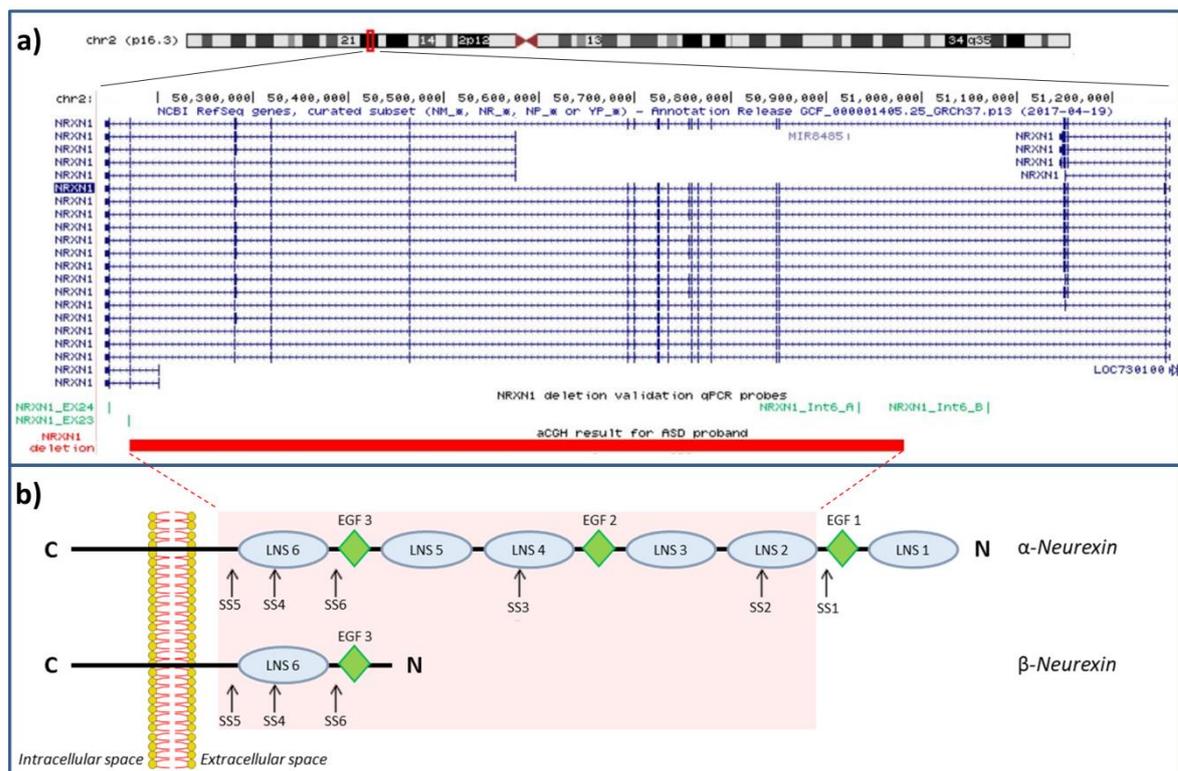


Figure 1. *NRXN1* deletion in the ASD proband. a) UCSC hg19 screenshot showing the *NRXN1* maternally inherited deletion detected by array-CGH in the female proband and validated by qPCR in the family. qPCR probes used for validation and inheritance testing are shown in green. b) Schematic representation outlining domains structure of α -neurexin and β -neurexin protein variants. Canonical splice site (SS) of neurexins are indicated by arrows. Protein region affected by deletion is highlighted in red.

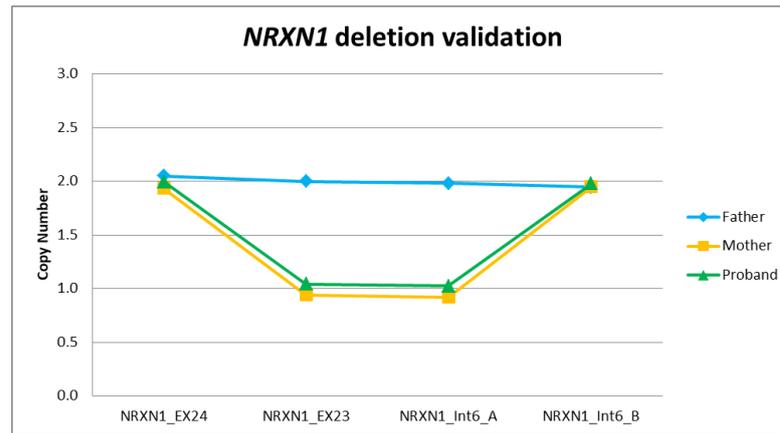


Figure 2. Results of validation and segregation analysis of *NRXN1* deletion.

To further explore the impact of rare variants background on the penetrance of *NRXN1* deletion, WES data of this family was analysed and filtered as illustrated in section 3.3.1. We focused our analysis on rare variants ($MAF \leq 0.5\%$) predicted to have a functional effect, including likely gene-disrupting (LGD) variants and missense variants defined damaging, according to a combination of prediction algorithms, as described by Yuen and colleagues (Yuen et al. 2015).

We hypothesized that development of the ASD phenotype in the proband and its absence in the *NRXN1* deletion-transmitting mother could be due to the higher risk genetic background of the proband, modulating the *NRXN1* deletion effect. Thus, we compared the load of rare variants in the proband and her mother, and we found a higher number of rare variants in the proband compared with the *NRXN1* deletion-transmitting mother (1036 versus 573, chi-squared $p = 8.05 \times 10^{-31}$). This difference remains significant by considering only the putative damaging variants, LGD and missense predicted damaging variants (303 in the proband vs 212 in the mother, $\chi^2 = 16.08$, $p = 6.07 \times 10^{-5}$). Then, we tested for transmission disequilibrium of damaging variants from the parents to the proband, detecting a preferential transmission of damaging variants from the father (203 transmitted vs 165 untransmitted variants, TDT $p = 0.048$) but not from the mother.

Moreover, we analysed genes with at least one LGD or putative damaging missense variant using the STRING database (Szklarczyk et al. 2019) to test the presence of an enrichment for functionally related networks of genes (Figure 3). This analysis identified a significant 1.2-fold enrichment (223 edges vs 184 expected) in interactions among the 294 genes identified in the proband ($p = 0.003$, one-tailed hypergeometric test), while no significant interaction enrichment was identified in *NRXN1* deletion-transmitting unaffected mother (84 edges vs 86 expected, $p = 0.62$).

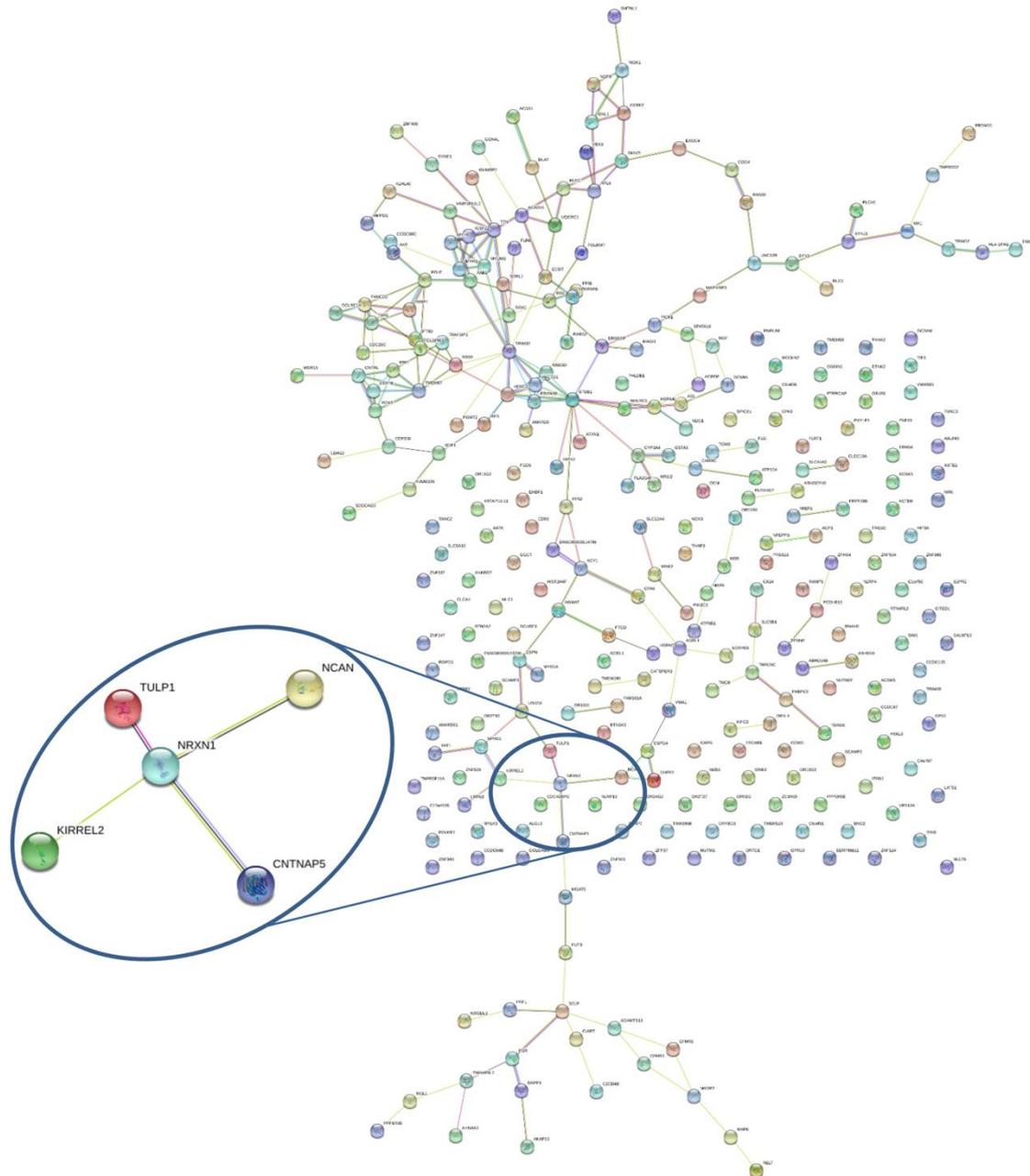


Figure 3. STRING network of predicted protein-protein interactions for genes harbouring LGD or likely damaging missense variants identified by WES in the female proband. The network of predicted *NRXN1* associations has been magnified. The edges represent the predicted functional associations and line colour indicates the type of interaction evidence: red line – presence of fusion evidence; green line – neighbourhood evidence; blue line – co-occurrence evidence; purple line – experimental evidence; yellow line – text-mining evidence; light blue line – database evidence; black line – co-expression evidence. *NRXN1* and *CNTNAP5* show homology, co-expression and text-mining interaction evidence; *NRXN1* and *KIRREL2* show text-mining interaction evidence; *NRXN1* and *NCAN* show co-expression and text-mining interaction evidence; *NRXN1* and *TULP1* show co-expression and experimental interaction evidence.

Next, we looked for specific sequence variants identified in the proband that might contribute to the clinical manifestation of the *NRXN1* deletion in the affected girl compared with the unaffected carrier mother. Firstly, we tested the possible presence of compound heterozygosity in *NRXN1*, but we found no rare functional nucleotide variants on the non-deleted allele. Then, we focused on four categories of rare variants: (a) likely damaging *de novo* variants, (b) recessive-acting variants (homozygous and compound heterozygous variants), (c) likely deleterious variants in genes previously implicated in ASD (reported in SFARI Gene database, <https://gene.sfari.org/>) that are intolerant to functional variation, having RVIS $\leq 20^{\text{th}}$ percentile (Petrovski et al. 2013) and/or pLI score ≥ 0.9 (Lek et al. 2016) and (d) likely deleterious variants in genes with a predicted interaction with *NRXN1* in STRING. We identified (Table 13):

- a) 1 *de novo* stop-gain in *CDC25C* (NP_073720.1:p.Ser143Xaa) and 1 *de novo* missense in *WASHC5* (NP_001317538.1:p.Asp254Gly);
- b) 2 homozygous missense variants in *ZFP37* and *DCLRE1A* and 2 compound heterozygous variants in *IFT80*;
- c) 5 missense variants in SFARI ASD candidate genes intolerant to mutation (*CNTNAP5*, *ERBIN*, *SYNE*, *HERC2* and *TANC2*). All of them, except for *TANC2*, were of paternal origin in brain expressed genes;
- d) 4 damaging missense variants in four genes that are predicted to interact with *NRXN1* (*CNTNAP5*, *TULP1*, *NCAN* and *KIRREL2*) (Figure 3). Among them, the missense variant in *CNTNAP5* is likely to exert a significant role, given that *CNTNAP5* is itself a previously known ASD candidate gene, likely to be intolerant to mutations (RVIS percentile = 8.4).

Finally, we investigated the mother's burden of rare variants in these risk categories and we detected a higher number of variants in the proband compared with the mother (11 vs 6 variants).

To have a more in-depth genetic characterization of this family, deep sequencing of the entire mtDNA was carried out in the ASD proband and her parents by the research group directed by prof. Valerio Carelli of the University of Bologna. No pathogenic variants were identified in the mtDNA of the proband. Among variants with a low-level heteroplasmy (between 0.2% and 15%), two maternally inherited variants were found in the proband with higher heteroplasmy (13.6% and 12.7%, respectively) compared with the mother (2.2% and 2.1%, respectively), while other five variants with a very low-level of heteroplasmy (0.2%-0.7%) were found exclusively in the proband (Cameli et al. 2021).

Table 13. Rare LGD and predicted damaging missense variants in four risk categories identified in the ASD proband. *pLI score ≥ 0.9 and/or RVIS percentile ≤ 20 are underlined; # predicted protein-protein interaction in STRING; ° PSD/Synaptic/Brain expressed gene according to (Bayés et al. 2011, Jansen et al. 2017, Kang et al. 2011); § brain expressed according to (Han et al. 2010); variants validated by Sanger sequencing are highlighted in bold.

Identified Variants							pLI score	RVIS percentile	PSD/Synaptic/Brain expressed°
Gene Base Change (hg19)	Amino Acid Change	Effect	Inheritance	dbSNP	MAF in gnomAD (exome)	Gene (SFARI score)			
<i>a) De novo variants</i>									
NC_000005.9:g.137627774G>T	NP_073720.1:p.(Ser143Xaa)	Stopgain	<i>De Novo</i>			<i>CDC25C</i>	3.01E-08	46.96	Brain expressed
NC_000008.10:g.126079907T>C	NP_001317538.1:p.(Asp254Gly)	Missense	<i>De Novo</i>			<i>WASHC5</i>	4.81E-13	<u>1.94</u>	PSD/ Brain expressed
<i>b) Homozygous and compound heterozygous variants</i>									
NC_000003.11:g.160000321A>T	NP_001177171.1:p.(Asp350Glu)	Missense	Maternal			<i>IFT80</i>	1.44E-10	50.94	Brain expressed
NC_000003.11:g.160037568T>C	NP_001177171.1:p.(Thr176Ala)	Missense	Paternal	rs146065418	0.0004781	<i>IFT80</i>	1.44E-10	50.94	Brain expressed
NC_000009.11:g.115812140G>T	NP_001269444.1:p.(Leu64Met)	Missense	Paternal/ Maternal	rs151180938	0.0029	<i>ZFP37</i>	4.11E-09	45.02	Brain expressed
NC_000010.10:g.115602192T>A	NP_055696.3:p.(Ile859Phe)	Missense	Paternal/ Maternal	rs11196530	0.002732	<i>DCLRE1A</i>	4.68E-08	70.92	Brain expressed
<i>c) Variants in SFARI Gene intolerant to mutation*</i>									
NC_000002.11:g.125660581G>A	NP_570129.1:p.(Ala1186Thr)	Missense	Paternal	rs114400050	0.00231	<i>CNTNAP5</i> (3)	0.1	<u>8.42</u>	Brain expressed
NC_000005.9:g.65321695C>G	NP_001006600.1:p.(Leu304Val)	Missense	Paternal	rs148121803	0.0001956	<i>ERBIN</i> (2)	<u>0.99</u>	63.5	PSD/ Synaptic/ Brain expressed
NC_000006.11:g.152642393C>G	NP_149062.1:p.(Asp5335His)	Missense	Paternal			<i>SYNE1</i> (3-S)	3.75E-27	<u>7.64</u>	PSD/ Brain expressed
NC_000015.9:g.28479425G>A	NP_004658.3:p.(Leu1337Phe)	Missense	Paternal	rs145594989	0.002391	<i>HERC2</i> (S)	<u>1</u>	<u>0.16</u>	Brain expressed

NC_000017.10:g.61498241A>G	NP_079461.2:p.(Tyr1633Cys)	Missense	Maternal	rs765761955	0.000004011	<i>TANC2</i> (1)	1	<u>0.33</u>	Brain expressed [§]
d) Variants in genes interacting with <i>NRXN1</i>[#]									
NC_000002.11:g.125660581G>A	NP_570129.1:p.(Ala1186Thr)	Missense	Paternal	rs114400050	0.00231	<i>CNTNAP5</i> (3)	0.1	<u>8.42</u>	Brain expressed
NC_000006.11:g.3546778T>C	NP_001276324.1:p.(Lys436Arg)	Missense	Maternal	rs62636511	0.00001989	<i>TULP1</i>	0.81	65.05	
NC_000019.9:g.19335134C>T	NP_004377.2:p.(Arg224Cys)	Missense	Maternal			<i>NCAN</i>	0.15	<u>6.15</u>	PSD/ Synaptic/ Brain expressed
NC_000019.9:g.36351545C>A	NP_001316459.1:p.(Arg252Ser)	Missense	Paternal	rs73928337	0.001664	<i>KIRREL2</i>	8.95E-06	93.31	Brain expressed

5. Results – Part II: Pipeline development for NGS data filtering and prioritization and *de novo* variants analysis

To further explore the role of rare variants in ASD susceptibility, provide additional support to genes already associated to ASD and identify new candidate genes, we performed a comprehensive analysis of the genomic variation of a clinically well-defined cohort of 105 families with at least one individual affected by ASD, by using WGS. This collection included 21 multiplex and 84 simplex families, and consisted of 124 ASD individuals, 210 parents and 58 unaffected siblings. Families recruitment was carried out by UOSI Disturbi dello Spettro Autistico, IRCCS Istituto delle Scienze Neurologiche of Bologna, Italy.

To manage the huge quantity of variants emerging from WGS and properly analyse them for the identification of putative causative variants, a semi-automated customised pipeline was developed for variants filtering and prioritization. Indeed, variants needed to be filtered to keep only rare coding and splicing variants, as coding variants are easier to interpret and more likely to be causative compared to non-coding variants, and prioritized according to several criteria. Our aim was to create a solid pipeline for variants selection, that could be also made automatic in order to have a reliable and efficient tool for NGS data analysis.

We had previously conceived and employed a manual pipeline for the analysis of WES data from the *NRXNI* deletion carriers family. Starting from that procedure, each step was updated and improved to better deal with larger and more complex data sets. Moreover, reference databases were carefully chosen to take into account recent updates and larger data sets, filtering parameters were carefully set to only include the variants that are more likely to have a deleterious effect, and prioritization criteria were wisely defined to select a numerically analysable set of variants without omitting any relevant one. Thus, we develop a robust and flexible pipeline, that could be suitable for any type of NGS data, from both WGS and WES. The developed pipeline was tested for the analysis of several families, then it was implemented in R, and used for the WGS data analysis of the entire cohort of 105 ASD families. A detailed description of the developed pipeline, whose schematisation is illustrated in figure 4, and of the WGS data analysis performed is provided hereafter.

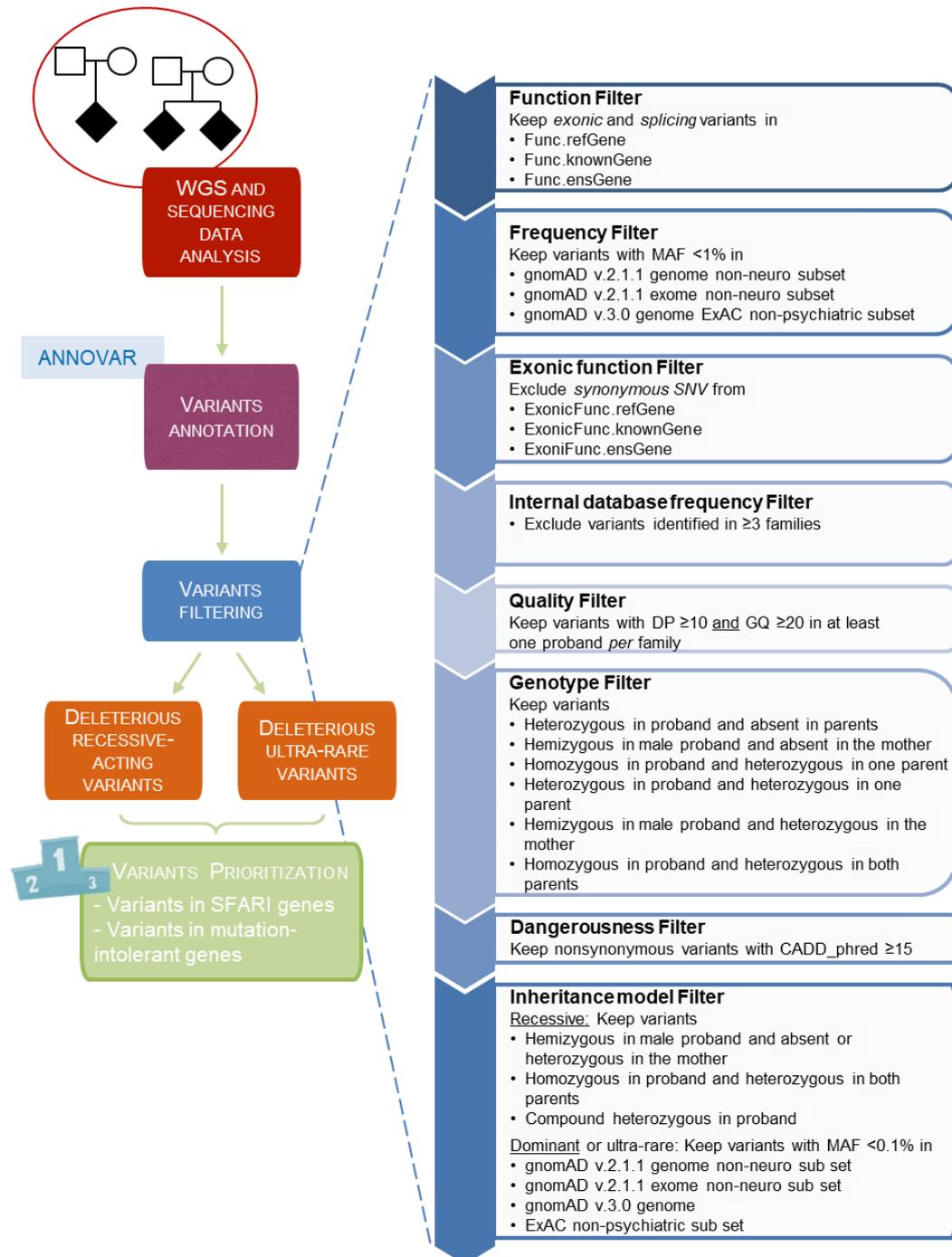


Figure 4. Schematisation of developed pipeline for WGS data filtering and prioritization.

After a first-level analysis of sequencing data, including quality check, reads trimming and mapping step, a vcf file was generated for each family. This contained all the variants (both Single Nucleotide Variants – SNVs – and small insertions and deletions – InDels) identified in a family. The variants were annotated using ANNOVAR software, using the latest genome build GRCh38/hg38. Then, annotated variants underwent the first filtering step: in order to select variants located in the coding sequence or affecting canonical splice site,

variants annotated as exonic or splicing were extracted from annotated files, generating a new data set to be further filtered and prioritized.

In the second filtering step, variants were filtered according to their Minor Allele Frequency (MAF) in the general population, which was estimated from public databases. Specifically, variants with MAF >1% in Genome Aggregation Database (gnomAD) and Exome Aggregation Consortium (ExAC) database (<https://gnomad.broadinstitute.org/>) (Karczewski et al. 2020) were removed. This allowed us to select only rare variants (MAF \leq 1%). Currently, gnomAD is the largest and most updated database of exome and genome sequencing data from a variety of large-scale sequencing projects, taking into account allele frequencies in different world-wide population and particular subsets. To avoid frequency bias, we used the non-neurological subset of both ExAC and gnomAD v.2.1, and the entire data set of gnomAD v.3.0. Both gnomAD versions were adopted in order to consider the most inclusive data set but sparing conversion bias, as at the time of analysing v.2.1 was the largest gnomAD data set but using GRCh37/hg19 for alignment, instead the more limited v.3.0 was built on GRCh38/hg38.

Subsequently, variants were filtered according to exonic function: by removing synonymous variants, a set of stop-gain, start-loss, frameshift, splicing and nonsynonymous variants was selected, as our aim was to identify and analyse variants with negative impact on protein, influencing protein sequence or altering normal splicing procedure.

In order to identify and remove artifacts specific to the sequencing protocols or downstream analysis pipeline, we developed an internal panel of “recurrent variants” against which NGS variants retained up to this point were filtered, to remove variants identified in 3 or more families of our cohort. This internal database is updated as families analysis proceeds, and represents a comprehensive variants repository, since it contains rare coding and splicing variants from all the families cohort.

A quality filter, based on genotype-level quality metrics, was applied to keep only variants with reliable call, hence having depth of coverage (DP) value \geq 10 and genotype quality (GQ) value \geq 20 in at least one proband *per* family.

The next step consisted of genotype filtering. VCF files emerging from sequencing data, and consequently annotated files too, contained variants identified in the entire analysed family, including variants absent in ASD individuals. To examine only variants that could be implicated in the disease development, a genotype-based filter was applied, and variants present in unaffected parents and/or siblings but not shared with affected subjects were

removed, along with variants for which one or both parents were homozygous while the proband was heterozygous.

As last filtering step, exonic variants were filtered according to dangerousness of their consequence on protein function: variants annotated as nonsynonymous were filtered by Combined Annotation-Dependent Depletion (CADD) score, keeping variants with CADD phred score value ≥ 15 and defining them as damaging missense variants. CADD score is one of the most used composite metric for the prioritization of genetic variants (both SNVs and InDels) according to their predicted deleteriousness. Through a machine learning model, CADD integrates annotation regarding conservation, epigenetic modification, functional prediction and genetic context into a single score predicting deleteriousness and potentially pathogenicity for each variant (Kircher et al. 2014). In our analyses we referred to both damaging missense and likely gene-disrupting (LGD) variants as deleterious variants. LGD variants, including stop-gain, stop-loss, frameshift and splicing variants, were kept regardless of CADD, even if they generally have high CADD score values. Indeed, by introducing a premature stop codon, altering the transcriptional frame or affecting the canonical splicing mechanism (by altering the two splice-site nucleotides to the left and right of each exon), LGD variants have a dramatic impact on the protein function and are considered to be the most deleterious variants.

Filtering steps here described allowed us to select for each family a stringent set of inherited and *de novo* rare deleterious coding and splicing variants, among which we distinguished two categories: rare (MAF $< 1\%$) recessive-acting variants and ultra-rare variants (MAF $< 0.1\%$). Recessive variants group consisted of variants acting under a recessive inheritance model, including homozygous, hemizygous and compound heterozygous (or biallelic) variants in probands. Conversely, dominant or ultra-rare variants group comprised more rare variants able to impact phenotype even affecting one gene allele but with incomplete penetrance. This subset was obtained by further filtering variants according to their MAF in the same previously used databases, and retaining only variants having MAF $\leq 0.1\%$.

To investigate the involvement of known ASD genes, recessive and ultra-rare variants were prioritized evaluating the affected gene and using SFARI Gene database and its scoring system to define genes previously implicated in ASD susceptibility (<https://gene.sfari.org/>). The Simons Foundation Autism Research Initiative or SFARI Gene database is a curated list of genes implicated in ASD susceptibility and classified into 4 categories based on the

evidence supporting their link to ASD: S (syndromic), 1 (high confidence), 2 (strong candidate) and 3 (suggestive evidence) (<https://gene.sfari.org/>, Release: 2021 Q3).

Moreover, deleterious variants (LGD and damaging missense variants) were prioritized on the basis of intolerance to mutations of the affected gene, adopting RVIS, pLI and LOEUF intolerance scoring systems. Residual variance to intolerance score (RVIS) assesses gene intolerance to functional variants ($RVIS \leq 20^{\text{th}}$ percentile, (Petrovski et al. 2013)), while genes with $pLI \geq 0.9$ (Lek et al. 2016) and $LOEUF \leq 0.35$ (Karczewski et al. 2020) are genes with strong probability of being intolerant to loss-of-function (LoF) mutations.

The data set of ultra-rare and recessive-acting variants generated by the application of the developed filtering and prioritization pipeline was employed to highlight variants with a potential causative role in ASD aetiology. Specifically, within the present project the variants data set was used to perform a preliminary analysis of *de novo* variants, and to explore the contribute of specific candidate genes.

5.1 *De novo* variants analysis

Among variants from WGS analysis we identified 223 ultra-rare (MAF <0.1%) coding and splicing deleterious variants appearing *de novo* in ASD individuals.

Given that the *de novo* false-positive call probability is generally high even using a highly precise variant calling tool, these variants were further analysed and filtered. First, all variants predicted as *de novo* were confirmed by visual inspection of the sequence alignments (BAM or CRAM files) by using the Integrative Genomics Viewer (IGV) software. This allowed us to exclude false positive cases, consisting of both artefactual calls in the proband and inherited variants under-called in one or both parents. Several variants with uncertain classification after this review in IGV were verified by Sanger sequencing, and resulted to be inherited or absent also in the proband.

Thus, by removing false positive variants, we obtained a set of 101 *de novo* variants, deriving from 75 ASD individuals belonging to 67 families (52 simplex and 15 multiplex families).

Among these variants, 14 *de novo* variants were located in known ASD gene (SFARI genes) and were validated by Sanger sequencing.

The set of 101 *de novo* variants consisted of:

- 23 LGD and nonsynonymous variants with CADD score value ≥ 30 ;
- 59 nonsynonymous variants with CADD score value < 30 and ≥ 20 ;
- 19 nonsynonymous variants with CADD score value < 20 and ≥ 15 .

The subset of LGD and missense variants with CADD score ≥ 30 includes the *de novo* variants that are predicted to have a stronger impact on the protein function and thus more likely to be able to explain the disease outcome (Table 14). Among them, we identified:

- 16 LGD variants, including 6 variants in genes predicted to be highly intolerant to loss-of-function (LoF) mutations, with pLi score ≥ 0.9 and LOEUF score ≤ 0.35 . Among LoF intolerant genes with LGD *de novo* variants, 2 genes (*SCN3A* and *IRX5*) were not previously associated with ASD, while 4 genes (*SHANK3*, *SCN2A*, *UBR5* and *BRSK2*) were also reported to be strong ASD candidate genes (SFARI gene with SFARI score 1 or 2)
- 7 missense variants with CADD score ≥ 30 , including 4 variants in genes not previously associated with ASD and 3 variants in SFARI genes (SFARI score 1, 2 or S).

Table 14. *De novo* LGD and CADD ≥ 30 variants identified from WGS in 105 ASD families. [§]pLi score ≥ 0.9 , LOEUF ≤ 0.35 and/or RVIS percentile ≤ 20 are underlined; novel variants are indicated in bold; all variants located in SFARI genes were validated by Sanger sequencing.

Genomic change (hg38)	Amino acid change	Effect (CADD score)	Gene (SFARI score)	pLi score [§]	LOEUF score [§]	RVIS percentile [§]	Proband ID
NC_000001.11: g.111695356C>T	NP_001010935.1: p.(Gln25Ter)	stopgain (40)	RAP1A	0.62	0.54	40.71	69.3
NC_000001.11: g.178457852C>T	NP_004832.1: p.(Gln713Ter)	stopgain (38)	RASAL2	0.8	<u>0.34</u>	<u>16.95</u>	19.3
NC_000002.12: g.165091265G>A	NP_008853.3: p.(Arg1630Ter)	stopgain (36)	SCN3A	<u>1</u>	<u>0.17</u>	<u>1.4</u>	40.3
NC_000002.12: g.165374892C>T	NP_066287.2: p.(Gln1394Ter)	stopgain (36)	SCN2A (1)	<u>1</u>	<u>0.13</u>	<u>1.62</u>	113.4
NC_000002.12: g.196317282A>G	NP_065811.1: p.(Leu809Pro)	missense (31)	HECW2 (2)	<u>1</u>	<u>0.26</u>	20.05	68.3
NC_000002.12: g.37975263_37975264 del	NP_653314.3: p.(Met405Valfs*2)	frameshift deletion	RMDN2	0	1.56	67.27	85.4
NC_000003.12: g.114082726_11408272 9del	NP_078914.1: p.(Lys318Leufs*10)	frameshift deletion	QTRT2	0	1.17	35.04	80.3
NC_000004.12: g.10445801_10445802 del	NP_444270.2: p.(Glu176Valfs*17)	frameshift deletion	ZNF518B	0	0.58	60.54	10.3
NC_000007.14: g.128672454_12867245 5del	NP_001012457.3: p.(Glu19Aspfs*27)	frameshift deletion	FAM71F2	0.02	1	80.18	95.3- 95.4
NC_000007.14: g.48108248G>A	NP_003355.1: p.(Arg275His)	missense (32)	UPP1	0	1.12	35.87	83.3
NC_000008.11: g.102265260del	NP_056986.2: p.(His2481Metfs*7)	frameshift deletion	UBR5 (2)	<u>1</u>	<u>0.07</u>	<u>0.2</u>	110.3

NC_000009.12: g.132904429C>A	NP_000359.1: p.(Asp675Tyr)	missense (31)	TSC1 (1)	<u>1</u>	<u>0.12</u>	<u>12.59</u>	100.3
NC_000010.11: g.89758704A>G		splicing (34)	KIF20B	0	0.69	79.73	67.3
NC_000011.10: g.1454559_1454560del	NP_001243556.1: p.(Asp540Glufs*9)	frameshift deletion	BRSK2 (1)	<u>0.95</u>	<u>0.33</u>	<u>4.94</u>	32.3
NC_000012.12: g.29514527del	NP_001180380.1: p.(Gly796Glufs*4)	frameshift deletion	TMTC1	0	0.72	44.04	112.3
NC_000013.11: g.41192409G>A	NP_115514.2: p.(Arg617Cys)	missense (31)	KBTBD7	0	0.92	<u>11.47</u>	40.4
NC_000014.9: g.100381472_10038149 0insGCGGCAGGCA TTAAGCACG	NP_078791.3: p.(Glu190Alafs*12)	frameshift insertion	WDR25	0	0.85	57.36	7.3
NC_000016.10: g.54931226C>T	NP_005844.4: p.(Gln10Ter)	stopgain (37)	IRX5	<u>0.97</u>	<u>0.3</u>	<u>16.34</u>	102.3
NC_000017.11: g.74777376G>A		splicing (33)	TMEM104	0	0.8	89.91	110.3
NC_000017.11: g.76926616G>A	NP_653278.2: p.(Asp393Asn)	missense (32)	MGAT5B	0.01	0.47	32.84	15.3
NC_000019.10: g.13025217T>C	NP_001257972.1: p.(Leu83Pro)	missense (32)	NFIX (S)	<u>1</u>	<u>0.15</u>	24.09	81.3
NC_000019.10: g.8399920G>A	NP_004209.2: p.(Arg33His)	missense (32)	RAB11B	0.44	0.64	33.58	52.3
NC_000022.11: g.50730801C>A	NP_001358973.1: p.(Ser1624Ter)	stopgain (41)	SHANK3 (1)	<u>1</u>	<u>0.12</u>	.	29.3

6. Results – Part III: *CACNA1H* biallelic variants in two unrelated ASD families

Given the involvement of voltage-gated calcium channels (VGCCs or Ca_v) genes in ASD susceptibility, rare variants data set generated from WGS analysis was employed to examine the presence of deleterious variants in genes coding for the VGCCs $\alpha 1$ principal subunit (*CACNA1A*, *CACNA1B*, *CACNA1C*, *CACNA1D*, *CACNA1E*, *CACNA1F*, *CACNA1G*, *CACNA1H*, *CACNA1I*, *CACNA1S*) and genes coding for the auxiliary subunits (*CACNA2D1*, *CACNA2D2*, *CACNA2D3*, *CACNA2D4*, *CACNB1*, *CACNB2*, *CACNB3* and *CACNB4*), since auxiliary subunits have also been found to contribute to ASD development likely due to their regulation of channel biophysical properties and $\alpha 1$ subunit targeting at the cell membrane (Dolphin 2012). Within this subset of variants in VGCCs genes, we looked both for ultra-rare *de novo* variants and more common variants acting under a recessive inheritance model (homozygous, hemizygous and compound heterozygous variants).

We identified 53 ultra-rare deleterious variants in 17 VGCCs genes in 40 ASD individuals (Appendix: Table S1). These variants were all inherited from unaffected parents and did not include any *de novo* variants. Among recessive-acting variants, we identified 2 hemizygous variants in *CACNA1F* and 4 compound heterozygous variants in *CACNA1H* (Table 15). The presence of recessive-acting mutations in two VGCCs genes led us to evaluate the hypothesis that a recessive model could be a shared mechanism of action for variants affecting VGCCs. Thus, we decided to investigate the role of biallelic variants in *CACNA1H*, as this gene is more strongly implicated in ASD susceptibility, according to SFARI Gene database (*CACNA1H* SFARI score = 2 (strong candidate), <https://gene.sfari.org/database/human-gene/CACNA1H>; *CACNA1F* SFARI score = 3 (suggestive evidence), <https://gene.sfari.org/database/human-gene/CACNA1F>).

Biallelic variants in the *CACNA1H* gene were identified in 3 ASD individuals, belonging to two unrelated ASD families. Specifically, two *CACNA1H* SNVs, one inherited from the unaffected father and one inherited from the unaffected mother, were detected in two female monozygotic ASD twins (proband 22.3 and proband 22.4), while other two *CACNA1H* SNVs, one paternally and one maternally inherited again, were detected in the male proband of a non-related trio (proband 105.3) (Figure 5). Family 105 also included one elder unaffected brother (105.4), for which DNA was not available.

Table 15. Rare recessive-acting damaging variants in VGCCs genes identified in our WGS data set.

Proband ID (sex)	Gene (SFARI score)	Genomic change (hg38)	Amino acid change	CADD score	Inheritance	dbSNP	MAF (gnomAD v.3.0)
105.3 (M)	<i>CACNA1H</i> (2)	NC_000016.10: g.1204361A>T	NP_066921.2: p.(Lys785Met)	24.7	Maternal	rs28365117	0.003
		NC_000016.10: g.1205207C>T	NP_066921.2: p.(Pro849Ser)	18.22	Paternal	rs370675810	0.0000977
22.3-22.4 (F-F)	<i>CACNA1H</i> (2)	NC_000016.10: g.1220303C>T	NP_066921.2: p.(Pro2124Leu)	19.03	Maternal	rs372453886	0.00006978
		NC_000016.10: g.1220945C>T	NP_066921.2: p.(Ser2338Phe)	16.99	Paternal	rs757713867	0.00002792
5.3 (M)	<i>CACNA1F</i> (3)	NC_000023.11: g.49211360C>T	NP_005174.2: p.(Ala1419Thr)	18.34	Maternal	rs782741094	0.0002
112.3 (M)	<i>CACNA1F</i> (3)	NC_000023.11: g.49211983C>T	NP_005174.2: p.(Gly1350Ser)	24.7	Maternal	rs782780521	0.00002839

All identified variants were rare coding nonsynonymous variants predicted to be damaging for the normal protein function (CADD score ≥ 15) (Kircher et al. 2014) (Table 15). Proband 105.3 paternal and maternal variants are located within *CACNA1H* exons 10 and 11, respectively. The maternal substitution A>T causes the substitution of a lysine with methionine (NP_066921.2:p.Lys785Met) in the cytoplasmic loop linking protein domains I and II; the paternal change C>T affects the last amino acid residue of the transmembrane segment 2 of the protein domain II, leading to the substitution of proline 849 with a serine (NP_066921.2:p.Pro849Ser) (Figure 5). Proband 22.3-22.4 maternal and paternal C>T variants are both located in exon 35, the last exon of the gene, causing respectively the amino acid change Pro2124Leu and Ser2338Phe (NP_066921.2) in the cytoplasmic C-terminal region of the protein (Figure 5). The biallelic status of these variants was not shared with 22.5 unaffected sister, who inherited only the paternal mutation.

No clear pathogenic sequence variants were identified in probands of both families from WGS analysis. Among the ultra-rare (MAF $\leq 0.1\%$) variants of uncertain significance (VUS) emerged from WGS analysis in the two families, the most interesting variants were one *de novo* missense variant in *PRSS2* and 9 inherited missense variants predicted to be damaging in SFARI genes (SFARI score 2 and 3) in proband 105.3, and one *de novo* novel missense variant in *ARFGEF3* and 20 inherited ultra-rare potentially deleterious variants (2 LGD and 18 damaging missense variants) in SFARI genes (with SFARI score 1, 2 and 3) in probands 22.3 and 22.4 (Appendix: Table S2).

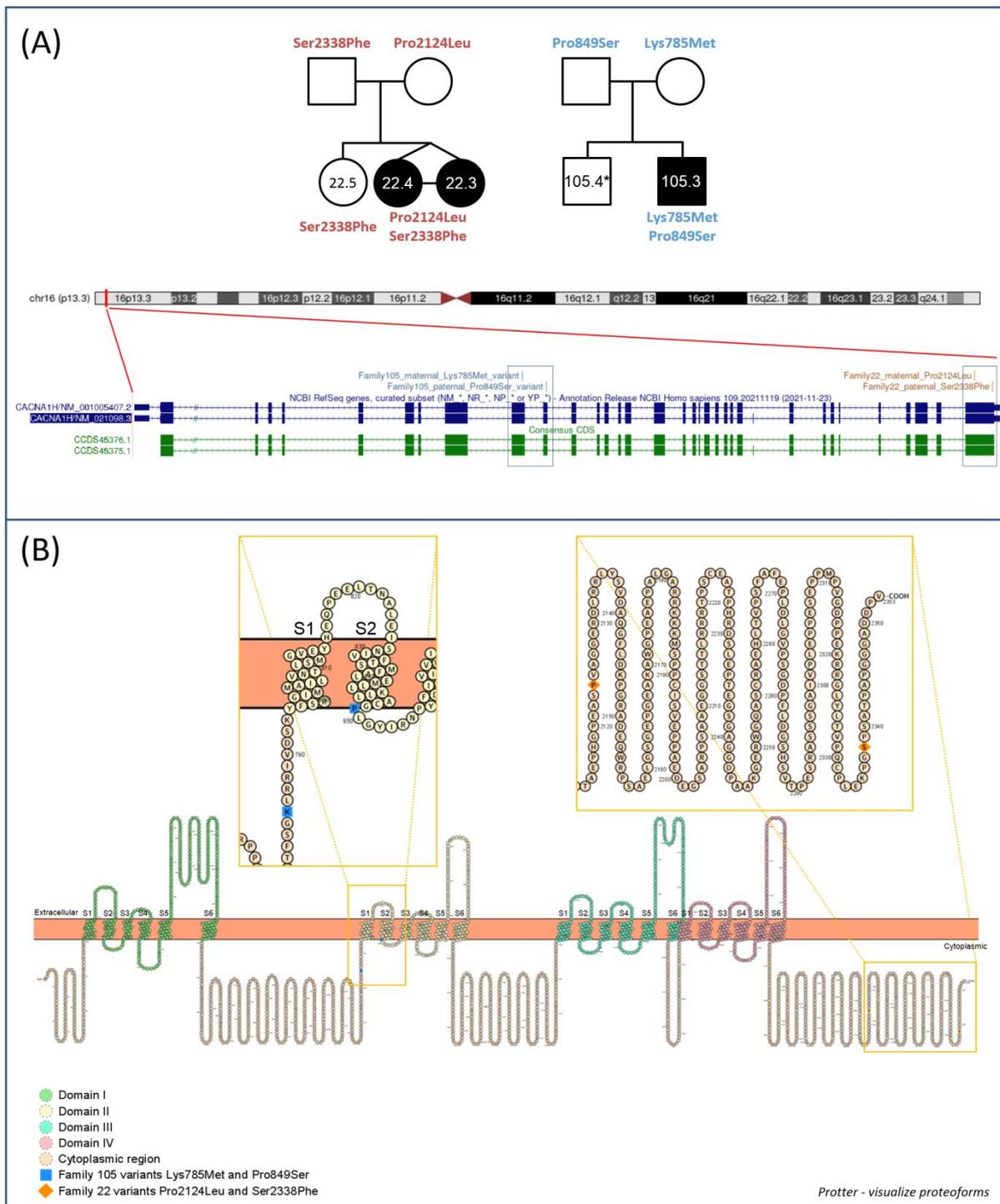


Figure 5. Family 22 and 105 *CACNA1H* variants. A) Family segregation and gene location of *CACNA1H* biallelic variants. Filled shapes indicate ASD individuals. *DNA was not available for individual 105.4. UCSC hg38 Genome Browser screenshot shows the location of biallelic variants within the *CACNA1H* gene. B) Schematic of the *CACNA1H* protein channel (Ca_v3.2). Ca_v3.2 channel consists of the single $\alpha 1$ pore-forming subunit of about 260 kDa, organized in four homologous domain each composed of six transmembrane segments (S1-S6). Within each domain, the arginine/lysine-rich S4 segment represents the voltage-sensing region of the channel, while the extracellular loop linking S5 and S6 segments (P loop) ensures the ion conductivity and selectivity of the channel (Perez-Reyes 2003, Talavera et al. 2001). Protein visualization was generated using Protter - visualize proteoforms (Omasits et al. 2014).

Heterozygous deleterious ultra-rare sequence variants in *CACNA1H* were found in other 9 ASD individuals of our cohort. However, these variants were all inherited from unaffected parents, and one of them was also shared with a unaffected sibling (Appendix: Table S1). No compound heterozygous condition for *CACNA1H* was identified in parents or unaffected siblings of our cohort.

In order to clarify the effect of the *CACNA1H* variants on the calcium channel activity, we generated 3xFLAG-Ca_v3.2 wild-type (WT) and 4 3xFLAG-Ca_v3.2 mutant recombinant proteins, corresponding to paternal and maternal mutations identified in families 22 and 105. Recombinant proteins were generated by cloning *CACNA1H* coding sequence (CDS) in the p3XFLAG-CMV-10 expression vector, to obtain an N-terminal 3xFLAG fusion protein detectable by anti-FLAG antibody and taking advantage of the increased detection sensitivity of this vector. Indeed, upstream of the multiple cloning region, p3XFLAG-CMV-10 contains three adjacent FLAG epitopes providing a considerable signal amplification of the recombinant protein. Moreover, the presence of the human cytomegalovirus promoter-regulatory region allows high vector replication efficiency in mammalian host-cells, like human embryonic kidney-293 (HEK-293T) cells, that were used for following experiments. *CACNA1H* coding sequence, derived from a1Ha-pcDNA3 plasmid (section 3.5), is about 7 kb and due to this large size, its cloning in the 3xFLAG vector was carried out in two different steps: firstly, 3'-*CACNA1H* CDS portion was cloned into an identically digested p3xFLAG-CMV-10; then, 5'-*CACNA1H* CDS portion was amplified by PCR reaction and inserted it in the previously obtained p3xFLAG-CMV-10-3'-*CACNA1H*. Thus, we generated p3xFLAG-CMV-10-CACNA1H-WT plasmid. To introduce the WGS identified mutations and obtain mutant plasmids, site-directed mutagenesis was performed on four different plasmids containing only a segment of *CACNA1H* CDS. Particularly, 5'- and 3'- *CACNA1H* CDS fragments were cloned in two p3xFLAG-CMV-10 vectors, that were used as template for four different whole-plasmid PCR reactions with divergent primers. Then, for each plasmid the missing *CACNA1H* CDS portion was added. In this way, we obtained four different p-3xFLAG-CMV-10-CACNA1H mutant plasmids, encoding for four different mutant recombinant proteins, all of them detectable by 3xFLAG (Figure 6). Sanger sequencing confirmed the absence of undesired changes within plasmids and the presence of desired mutations in mutant ones.

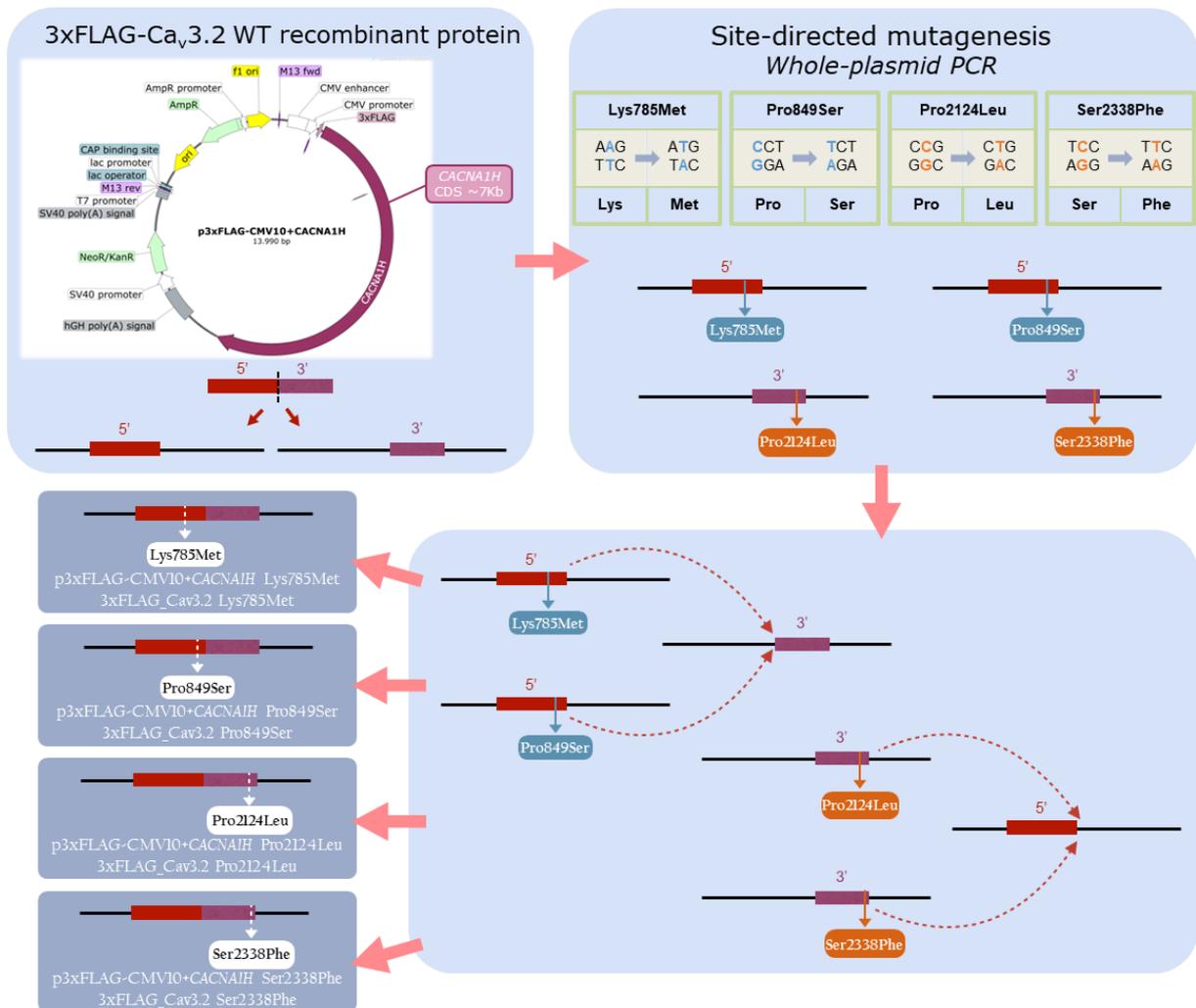


Figure 6. Schematic representation of molecular cloning strategy for the generation of WT and mutant Ca_v3.2 plasmid constructs.

In order to assess recombinant proteins expression in mammalian cells and verify their cellular localization, we performed immunofluorescence (IF) assay in HEK-293T cells, by transiently transfecting cells with p3xFLAG-CMV-10-CACNA1H WT and mutant plasmids. Moreover, an empty vector was used as negative control, while a 3xFLAG plasmid encoding the 3xFLAG-ABCC3 transmembrane fusion protein was used as positive control. Indeed, in our laboratory it is known that 3xFLAG-ABCC3 fusion protein cellular localization is not affected by 3xFLAG presence and that it correctly localizes at cell membrane.

As illustrated in figure 7, by merging Hoechst and Cy3 signals, indicating nuclei and recombinant proteins respectively, WT and mutant recombinant proteins were expressed in our cell system and correctly localize at the cell membrane, with no considerable differences in cellular localization among them and in comparison with the positive control (Figure 7).

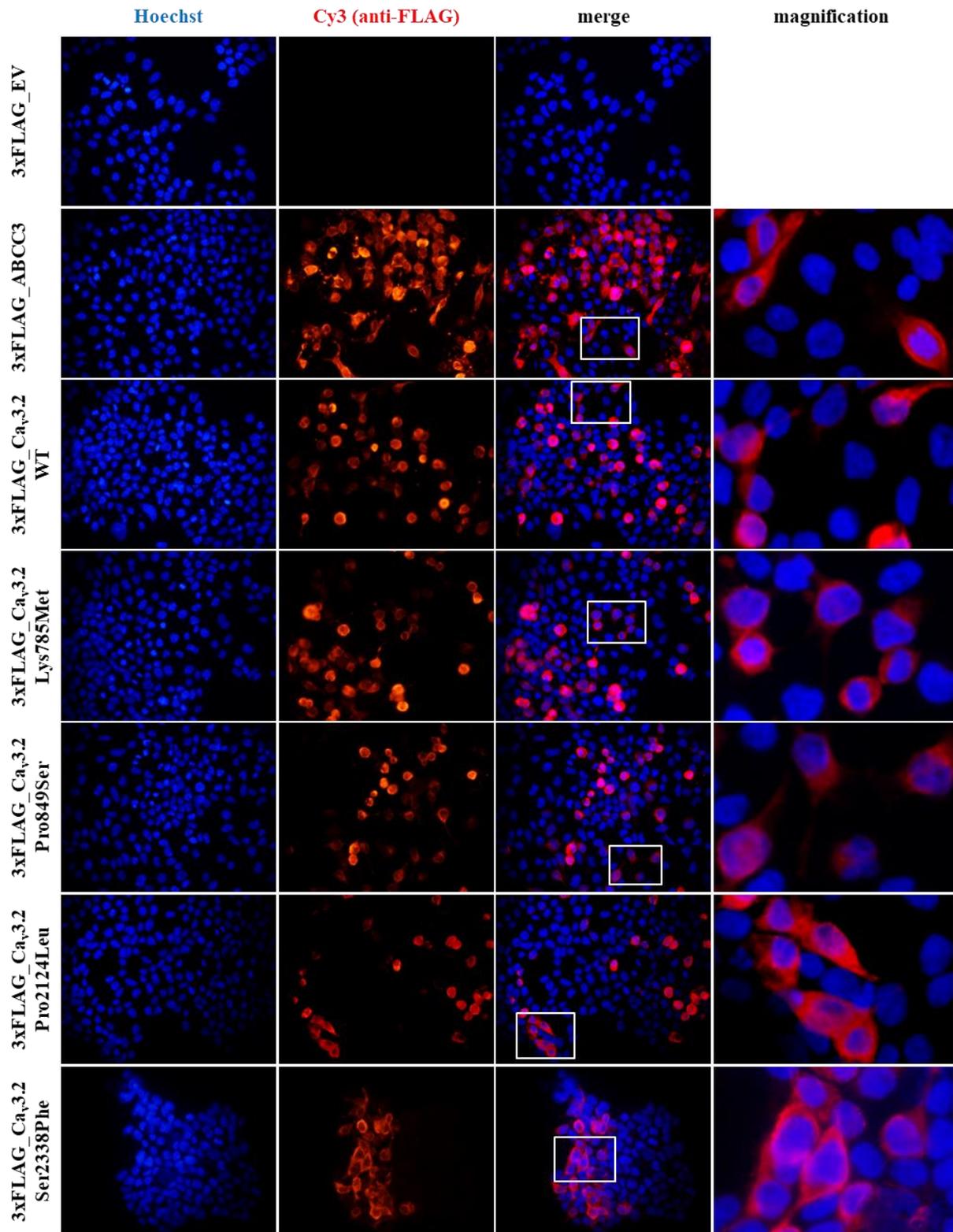


Figure 7. Immunofluorescence (IF) assay of 3xFLAG-Ca_v3.2. IF assay was performed in HEK-293T cells, by transiently transfecting cells with p3xFLAG-CACNA1H plasmid constructs. An empty vector (EV) was used as negative control, while 3xFLAG-ABCC3 transmembrane fusion protein was used as positive control. Mouse anti-FLAG M2 antibody was used for detection of the recombinant proteins. Goat anti-mouse Cy3 antibody (red signal) and Hoechst dye (blue signal) were employed to detect anti-FLAG antibody and nuclei respectively.

To unveil potential functional effects due to identified SNVs, electrophysiological experiments were then performed using the whole-cell patch-clamp technique, in collaboration with the research group directed by Prof. Sergio Fucile at the Department of Physiology and Pharmacology, Sapienza University, Rome. Plasmid encoding mutant 3xFLAG-Ca_v3.2, along with plasmid encoding the WT isoform, were transfected in HEK-293 cells, and the resulting voltage-activated Ca²⁺ currents were recorded and analysed (Figure 8). Cells transfected with plasmid DNA coding for the mutant proteins exhibited inward currents in response to depolarization steps, similarly to WT protein (Figure 8A, B). None of the mutant channels showed a loss-of-function effect, while current densities (Figure 8C, D) and activation properties (Figure 8E, F) were differently modulated by distinct variants. In cells transfected with plasmid DNA encoding the Lys785Met-mutant Ca_v3.2, we detected a larger mean current density (Figures 8D, 9A), and a left shift in the activation I-V curves, with a more hyperpolarized V_h value of -50 ± 1 mV, compared to the WT value of -45 ± 2 mV (Figure 9B; $p=0.037$). The inactivation properties were not affected by the mutations (Figure 8G, H).

The kinetics of inward currents, elicited by a -40 mV depolarization step, were differently affected by the mutations: compared with the WT values, currents mediated by the Ser2338Phe, Pro849Ser and Lys785Met mutant channels showed a slower time to peak (Figure 9C). Moreover, current mediated by the Pro849Ser isoform decayed more rapidly compared to the WT (Figure 9D). To evaluate the impact of these altered parameters on the channel function, we measured the total charge transfer (i.e., the total amount of Ca²⁺ ions entering the cells) induced by a -40 mV depolarization step, at different time points (2, 10 and 100 ms): a significant increase was observed at all time points only in cells transfected with the Lys785Met-mutant Ca_v3.2 plasmid construct (Figure 9E).

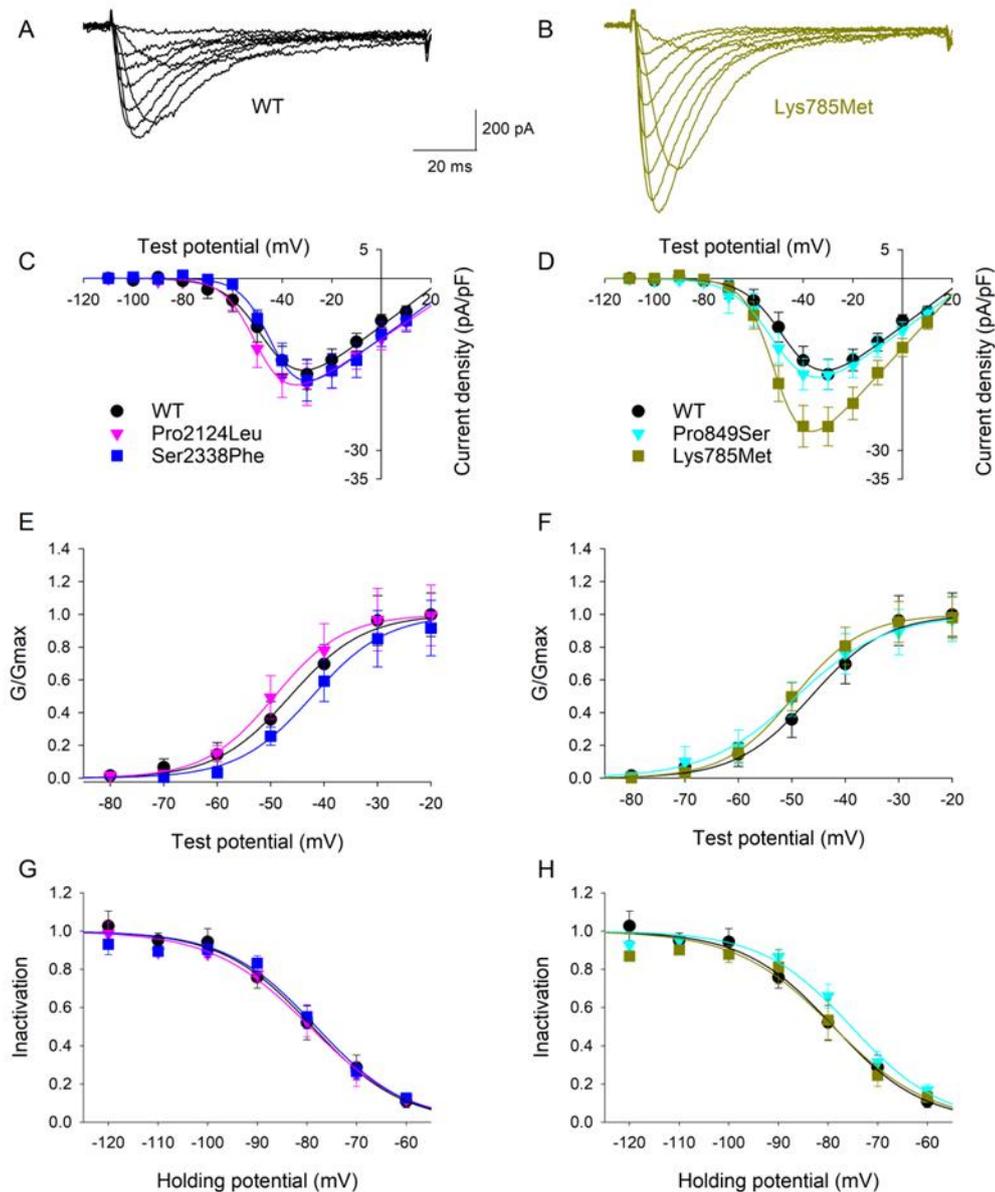


Figure 8. Functional characterization of WT and mutant $Ca_v3.2$ channels transiently expressed in HEK-293 cells. A) Representative current traces recorded from a HEK-293 cell transiently transfected with the WT CACNA1H plasmid construct. The currents were elicited by step depolarizations from a holding potential of -120 mV to various test potentials. B) Representative current traces recorded from a HEK-293 cell transiently transfected with the Lys785Met plasmid construct, same protocol as A). C) Mean activation curves for WT, Pro2124Leu and Ser2338Phe plasmid constructs transfected cells, as indicated. Solid lines represent data fit to the Boltzmann equation (V_h values are -45.5 mV, -49.3 mV and -43.4 mV for WT, Pro2124Leu and Ser2338Phe, respectively). Data were averaged from 16, 19 and 17 cells, for WT, Pro2124Leu and Ser2338Phe, respectively. D) Mean activation curves for WT (same data as C), Pro849Ser and Lys785Met plasmid constructs transfected cells, as indicated. Solid lines represent data fit to the Boltzmann equation (V_h values are -45.5 mV, -49.5 mV and -50.1 mV for WT, P849S and K785M, respectively). Data were averaged from 16 (same cells as C), 16 and 14 cells, for WT, Pro849Ser and Lys785Met, respectively. E) Normalized mean activation curves for WT, Pro2124Leu and Ser2338Phe plasmid constructs transfected cells, same data as C. Solid lines represent data fit to the activation Boltzmann equation. F) Normalized mean activation curves for WT, Pro849Ser and Lys785Met plasmid constructs transfected cells, same data as D. Solid lines represent data fit to the activation Boltzmann equation. G) Inactivation open probability-voltage relationships for WT, Pro2124Leu and Ser2338Phe plasmid constructs transfected cells. Solid lines represent data fit to the inactivation Boltzmann equation (V_i

values are -78.9 mV, -79.5 mV and -78.2 mV for WT, Pro2124Leu and Ser2338Phe, respectively). No significant differences were detected. H) Inactivation open probability-voltage relationships for WT, Pro849Ser and Lys785Met plasmid constructs transfected cells. Solid lines represent data fit to the inactivation Boltzmann equation (V_i values are -78.9 mV, -75.3 mV and -77.1 mV for WT, Pro849Ser and Lys785Met, respectively). No significant differences were detected.

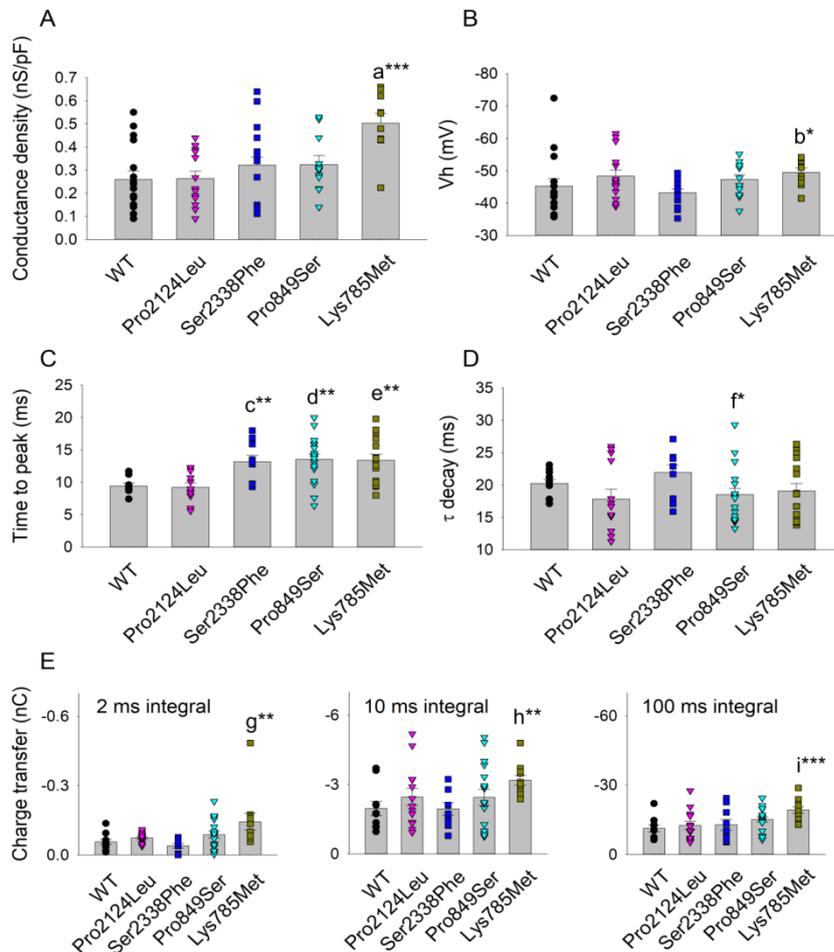


Figure 9. Activation and kinetic parameters of Ca^{2+} currents mediated by WT and mutant channels expressed in HEK-293 cells. A) Histogram representing the mean conductance density measured at -40 mV test potential for each $\text{Ca}_v3.2$ subtype, as indicated. Mean values were averaged from 16, 15, 13, 12 and 10 cells, from left to right. The conductance density was significantly higher for Lys785Met mutant, as compared to WT (a, $p < 0.001$). B) Histogram representing the mean V_h value measured at -40 mV test potential for each $\text{Ca}_v3.2$ subtype, as indicated. Same cells as A. The V_h value was significantly higher for Lys785Met 431 mutant, as compared to WT (b, $p = 0.037$). C) Left, histogram representing the mean values of time to peak measured at -40 mV test potential for each $\text{Ca}_v3.2$ subtype, as indicated. Mean values were averaged from 10, 13, 10, 18 and 15 cells, from left to right. The time to peak was significantly higher for Ser2338Phe, Pro849Ser and Lys785Met mutants, as compared to WT (c, $p = 0.002$; d, $p = 0.002$; e, $p = 0.003$). D) Histogram representing the mean values of exponential τ decay measured at -40 mV test potential for each $\text{Ca}_v3.2$ subtype, as indicated. Same cells as C. The time to peak was significantly higher for Pro849Ser mutant, as compared to WT (f, $p = 0.039$). E) Histograms representing the mean values of charge transfer measured at -40 mV test potential for each $\text{Ca}_v3.2$ subtype, as indicated, at different current times (left, 2 ms; center, 10 ms; right, 100 ms). Same cells as C. The charge transfer value was significantly higher only for Lys785Met mutant, at each time point, as compared to WT (g, $p = 0.004$; h, $p = 0.010$; i, $p < 0.001$).

7. Discussion

7.1 An increased burden of rare exonic variants in *NRXNI* microdeletion carriers is likely to enhance the penetrance for autism spectrum disorder

NRXNI encodes neurexin 1, a protein belonging to a family of presynaptic cell adhesion molecules forming a trans-synaptic complex with the postsynaptic adhesion proteins neuroligins, involved in the formation and maintenance of synaptic connections and vesicular neurotransmitter release.

CNVs affecting the *NRXNI* gene, specifically *NRXNI* deletions, have been widely associated with different neurodevelopmental and psychiatric disorders, including schizophrenia, ID, epilepsy, developmental disorder and ASD. Particularly, *NRXNI* is a strong candidate for ASD, as *NRXNI* deletions are among the most commonly reported rare genetic variants in individuals with ASD. However, these mutations often show incomplete penetrance, being inherited from unaffected or mildly affected parents, and absence of phenotype specificity, with the same genetic alteration leading to different clinical conditions in *NRXNI* deletion carriers. Consequently, determining the exact contribution of these mutations to the disease development is often difficult (Avazzadeh et al. 2019, Cosemans et al. 2020).

We identified a pathogenic *NRXNI* microdeletion in an ASD girl of our cohort, who inherited the CNV from the unaffected mother, as revealed by qPCR assay. Thus, we performed a characterization of rare variants both in the nuclear and mitochondrial genome of all family members, to explore the genetic background of the *NRXNI* microdeletion carriers and determine the most likely factors modulating the penetrance and expressivity of the *NRXNI* deletion.

The identified microdeletion overlaps *NRXNI* exons from 7 to 23 (NM_001135659.2), thus it can be classified as a 3' deletion (Cosemans et al. 2020), resulting in the loss of the major part of extracellular portion of both the two protein isoforms (from Gly311 to Leu1445). Indeed, in neurons the *NRXNI* gene is transcribed from two independent promoters, generating two different protein isoforms: a longer and more abundant in brain α -neurexin, and a shorter β -neurexin. Both of them are composed by a single transmembrane domain, with an identical intracellular sequence and distinct extracellular domains (Südhof 2017). The identified deletion gives rise to a putative in-frame transcript, lacking all α -neurexin LNS-domains (laminin/neurexin/sex hormone-binding globulin domains) except the first one, and the two

EGF-like domain, while maintaining the transmembrane and intracellular C-terminal domain (Figure 1, chapter 4). Moreover, the deletion affect the canonical splice sites (SS2 to SS6), including SS4 that is considered a key mechanism for the regulation of neurexin-ligand interactions at synapses (Südhof 2017). Since it is in-frame, it is likely that the deletion phenotypic effect may arise from two concurrent mechanisms: haploinsufficiency due to the absence of wild-type α -neurexin isoform, and a dominant-negative activity of the mutant splice isoform, as previously suggested by Flaherty and colleagues in their study on induced pluripotent stem cell (hiPSC)-derived neurons from heterozygous intragenic *NRXN1* deletion carriers (Flaherty et al. 2019).

The proband's phenotype was mainly compatible with 3'-*NRXN1* deletion clinical features, as in addition to ASD she exhibited macrocephaly and developmental disability, that are more frequently present in 3'deletion carriers in comparison with 5'deletion carriers (Cosemans et al. 2020). Instead, no broad autism phenotype traits were present in both parents, according to SCDC and BAPQ questionnaires scores (Skuse, Mandy and Scourfield 2005, Hurley et al. 2007).

From WES analysis carried out in all family members a higher number of rare coding variants emerged in the ASD girl in comparison with the *NRXN1* deletion-transmitting mother, and this difference remained statistically significant when the analysis was restricted to rare likely deleterious variants only. This suggested that these additional rare variants may contribute collectively to push the genetic liability beyond the threshold for ASD. Moreover, a significant interaction enrichment was detected among genes with damaging variants (CNV and SNVs) in the proband and not in the mother, supporting a cumulative effect of interacting genes affected by mutations to the phenotype.

Interestingly, four genes affected by damaging variants in the proband emerged to be predicted *NRXN1* interactors in the STRING analysis. Among them, the predicted interaction with *CNTNAP5* is of particular interest, since it is a functionally intolerant gene previously identified as a suggestive ASD candidate gene (SFARI score 3). *CNTNAP5* encodes for contactin-associated protein-like 5, a member of the neurexin family involved in cell adhesion and intercellular communication in the vertebrate nervous system (Traut et al. 2006). In ASD proband of our cohort, it harboured a paternally inherited putative damaging missense variant. Rare deletion and missense variants in *CNTNAP5* have been identified in subjects with ASD in a previous study of our group (Pagnamenta et al. 2010) and, similarly to our *NRXN1* microdeletion ASD carrier, Vaags and colleagues have previously identified a maternally inherited missense variant in *CNTNAP5* segregating with a *NRXN3* paternal deletion in two

ASD siblings (Vaags et al. 2012), supporting a combined role of neurexins and contactin-associated proteins in ASD risk.

In addition to *CNTNAP5*, damaging missense variants were identified in four other genes predicted to be intolerant to mutation and previously implicated in ASD: *TANC2*, *ERBIN*, *SYNE1* and *HERC2* (De Rubeis et al. 2014, Krumm et al. 2015, Iossifov et al. 2014, Liu et al. 2018, Guo et al. 2019, Yu et al. 2013, O'Roak et al. 2012, Wu et al. 2020, Li et al. 2014, Ruzzo et al. 2019, Stessman et al. 2017). *SYNE1* and *HERC2* have been mostly involved in syndromic forms of ASD, while *TANC2* and *ERBIN* are high confidence susceptibility genes for idiopathic ASD (SFARI score 1 and 2, respectively), encoding for postsynaptic proteins involved in synaptic processes regulation. Specifically, *TANC2* is highly expressed in the human developing brain and encodes for a postsynaptic scaffold protein involved in dendritic spines and in the regulation of excitatory synapses (Han et al. 2010). *ERBIN* postsynaptic protein binds ERBB2, playing an important role during brain development and regulation of synaptic plasticity in the adult brain (Huang et al. 2000). Moreover, it is also implicated in dendritic morphogenesis by regulating localization and function of δ -Catenin in hippocampal neurons (Arikkath et al. 2008).

From WES analysis two *de novo* novel putative damaging variants also emerged in the proband: a *de novo* stop-gain variant affecting exon 5 of the *CDC25C* gene (NP_073720.1:p.(Ser143Xaa)), and a missense variant predicted to be damaging occurring in *WASHC5* exon 9 (NP_001317538.1:p.(Asp254Gly)). Neither of the two genes have been previously implicated in ASD, however both of them are expressed in the brain (Kang et al. 2011) thus they could contribute to the proband phenotype. Furthermore, concerning *WASHC5*, a noteworthy aspect is that heterozygous missense variants in this gene are associated with autosomal dominant spastic paraplegia 8 (SPG8), a progressive upper- motor neurodegenerative disease (Ginanneschi et al. 2020), while biallelic pathogenic variants are associated with Ritscher-Schinzel Syndrome, characterized by distinctive craniofacial features, cerebellar defects and cardiovascular malformations (Elliott et al. 2013). It is not possible to rule out a pathological role for the identified *WASHC5* variant for SPG8 phenotype, which usually has on onset in adult life, and although the proband did not show typical dysmorphic craniofacial features Ritscher-Schinzel Syndrome, she presented macrocrania.

Recessive-acting variants were identified in three genes (*ZFP37*, *DCLRE1A* and *IFT80*). All of them are reported to be brain-expressed (Kang et al. 2011), but neither of them have been previously implicated in neurodevelopmental phenotypes.

In order to perform a more comprehensive analysis of the genomic variability in the *NRXNI* microdeletion ASD carrier family, rare pathogenic mutations in the mitochondrial genome and the burden of low-heteroplasmy variants were also assessed. Obtained results suggested that the burden of low-level heteroplasmic mtDNA variants might contribute to the ASD risk. However, further analyses on large cohorts are needed to validate this hypothesis.

Taken together, our results are consistent with a multi-factorial threshold model for ASD development in this family (Figure 10) and, more generally, in ASD cases with large-effect but incompletely penetrant susceptibility variants. The presence of an increased burden of exonic rare variants in the proband compared to the unaffected deletion-transmitting mother supports the hypothesis that the *NRXNI* deletion sensitizes the genome to a clinical manifestation, but other genetic contributors are necessary to cross the threshold for a phenotypic manifestation (Hoang, Cytrynbaum and Scherer 2018). Moreover, the reduced penetrance of the *NRXNI* deletion in the unaffected mother supports the female protective effect hypothesis, which suggests that females require an excess burden of deleterious CNVs and SNVs to reach the ASD diagnostic threshold (Jacquemont et al. 2014). The paternal-inherited rare variants in ASD-related or functionally constrained genes and *de novo* rare variants identified in the proband may have additive effects, acting on a sensitized background caused by haploinsufficiency and/or a dominant-negative activity at the *NRXNI* locus. This observation is in line with what has been already reported for neurodevelopmental phenotypes in probands with 16p21.1 deletions and in probands with other gene disruptive variants (Pizzo et al. 2019), and with the more general hypothesis that psychiatric phenotypes are modulated by multifactorial determinants even in the presence of a large-effect variants. Our results support the hypothesis that the burden of rare variants contributes in defining the phenotypic trajectory in carriers of a large-effect variant, modulating the penetrance and the expressivity of disease-associated variants. However, it has been shown that also common polygenic variation contributes additively to ASD risk, even in cases that carry a strongly acting variant (Robinson et al. 2014, Weiner et al. 2017). In the current study, the potential contribution of common variation was not possible to be evaluated, but we aim to conduct a complete assessment of the genomic landscape of the family through Polygenic Risk Score (PRS) evaluation, using SNP-array genotyping data that we are currently generating. Moreover, further investigation in a large data set will be necessary to properly evaluate the cumulative effects of rare deleterious and common variants to ASD risk in *NRXNI* deletion carriers.

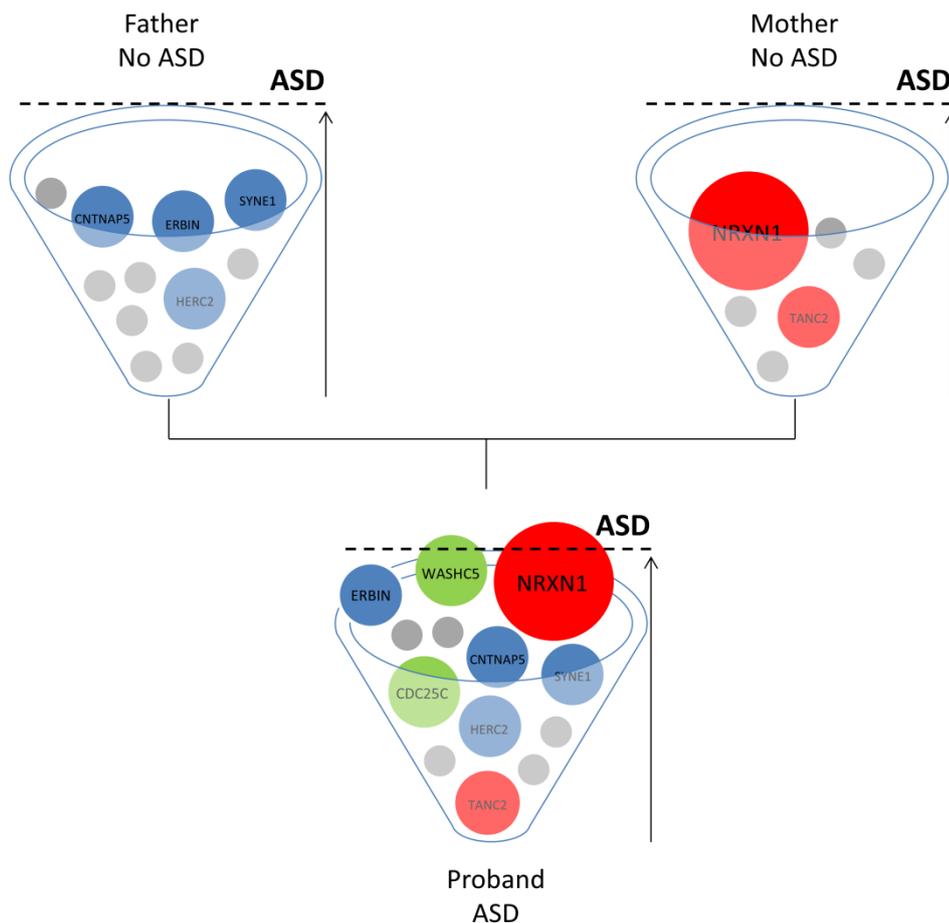


Figure 10. Multi-factorial threshold model for ASD in this family. Each family member has an ASD risk cup with balls representing risk variants that contribute to ASD with variable degrees of impact. In both parents, the burden of risk variants is not enough to develop ASD, while in the child the ASD threshold is reached as a combination of strong and weak, inherited and *de novo* genetic variants. The *NRXN1* deletion is depicted as a strong, primary contributing factor to reaching the ASD threshold in the ASD child, but not sufficient alone to develop ASD in the deletion carrier mother (Hoang et al. 2018)

7.2 *De novo* variants analysis

In order to better delineate the role of rare variants in ASD susceptibility and clarify the genetic architecture of the disorder, in the second part of this project we performed WGS analysis in a cohort of 105 ASD families. A reliable and efficient pipeline was developed for WGS variants filtering and prioritization, to advantageously manage and properly analyse generated data. The developed pipeline allowed us to select a stringent set of inherited and *de novo* variants, converging in two different groups: ultra-rare variants acting under a dominant inheritance model, and rare recessive-acting variants, including homozygous, hemizygous and biallelic variants.

Within the subset of ultra-rare variants we performed a preliminary analysis, focusing on *de novo* variants, that are more likely the leading ASD causes, while inherited variants are more likely to act as risk contributors in a cumulative model. Indeed, the leading role of *de novo* variants in the development of ASD and the significance of their analysis regarding the identification of candidate genes, and the interpretation of their role, is well established.

We explored the presence of ultra-rare *de novo* coding and splicing variants in our WGS data set. After filtering activity, according to variants frequency in the study sample, and results of sequence alignment visual inspection and Sanger sequencing of ambiguous cases, we identified a set of 101 *de novo* variants in 75 individuals with ASD. Within this group, variants prioritization was needed in order to aid the interpretation of variants role. Thus, variants were classified according to their impact on the normal protein function, expressed by variant location in the protein and the CADD score, to distinguish more clearly causative variants from variants with a likely causative or uncertain role in ASD development. Specifically, three categories of *de novo* variants were obtained: variants with most deleterious effect, consisting of likely gene-disrupting (LGD) and missense variants with CADD score ≥ 30 ; missense variants with CADD score < 30 and ≥ 20 ; variants with CADD score value < 20 and ≥ 15 .

For a first-level analysis, we explored the first group of *de novo* variants, including variants with the strongest impact on the normal protein function, thus being more likely involved in the definition of phenotypic manifestation, with the aim to identify genes of interest to follow-up with more in-depth analyses to clarify their involvement in the disorder. LGD and CADD ≥ 30 group includes 7 variants in strong ASD candidate genes (SFARI score 1, 2 or S) and 16 variants in genes not previously implicated in ASD.

Within SFARI genes, the major part consists of genes known to be clearly implicated in the development of ASD and whose *de novo* mutations are generally highly penetrant so that they are sufficient to explain ASD phenotype. This is the case of *de novo* variants in *SHANK3*, *SCN2A*, *TSC1* and *NFIX*. All these genes are clearly implicated in well-known syndromic forms of ASD, and their identification is of great interest for the clinical management of affected individuals and for a better characterization of syndromic ASD.

The other SFARI genes affected by LGD and CADD ≥ 30 *de novo* variants in our data set, *BRSK2*, *UBR5* and *HECW2*, are strong ASD candidate genes, as they have shown to harbour rare deleterious variants in ASD individuals. However, how they exactly contribute to ASD phenotype outcome is still poorly explored.

For example, to date 13 rare coding or splicing variants have been identified in the *BRSK2* gene in individuals with ASD or other neurodevelopmental disorder (NDD). The majority of these variants, emerged from large-scale sequencing studies, are *de novo* variants with a strong impact on the normal protein function, mainly acting as loss-of-function (Iossifov et al. 2014, De Rubeis et al. 2014, Hiatt et al. 2019, Feliciano et al. 2019). *BRSK2* encodes for a serine/threonine-protein kinase, playing a key role in neurons polarization and axonogenesis, cell cycle progress and insulin secretion, and it is a constraint gene, with a highly probability of being loss-of-function intolerant (Lek et al. 2016). Both the two *de novo* variants identified in *BRSK2* in the present study are of interest for further characterization analyses: the frameshift deletion, not included in the gnomAD database, is predicted as pathogenic (<http://varso.me/XrzT>), as the 2 bp deletion causes the formation of truncated protein isoform subjected to degradation by nonsense-mediated decay (<https://nmdprediction.shinyapps.io/nmdescpredictor/>) (Coban-Akdemir et al. 2018), and the nonsynonymous variant is predicted as likely pathogenic (<http://varso.me/Xm2e>) and affects the protein kinase domain. Therefore, *BRSK2* represents an interesting candidate gene for further analyses, aimed at deepening the role of the identified *de novo* variants on the protein, to understand whether they could alter pathways in which *BRSK2* is involved and how it could contribute to ASD development. A useful strategy to further explore the role of *BRSK2* in ASD could be to exploit platforms for sequencing data sharing to track down additional unpublished and recently identified *BRSK2* variants, to further describe the recently-emerged disease association maybe by a multidisciplinary approach on larger cohort.

The majority of *de novo* LGD and CADD ≥ 30 variants identified in our cohort are located in genes not clearly implicated in ASD, some of which appear to be promising gene to be further explored for involvement in ASD. A *de novo* stop-gain in one ASD proband might shed light on the role of the *SCN3A* gene in ASD. *SCN3A* encodes for the type-III voltage-gated sodium channel α -subunit $\text{Na}_v1.3$, belonging to the voltage-gated sodium channels (SCNs) family, that includes nine transmembrane tissue-specific proteins, consisting of an α subunit cooperating with one or more auxiliary β subunits. Analogously to VGCCs, the SCNs α subunits include four homologous domains, D1– D4, containing six α -helical transmembrane segments, S1-S6. Segment S4 harbours the voltage sensor, while S5 and S6 are the pore-forming segments: S5 and S6 segments constitute the inner pore and the S5-S6 extracellular loop constitutes the ion selectivity filter (Catterall 2000). The *de novo* variant identified in one ASD individual of our cohort is a novel predicted pathogenic stop-gain, considerably affecting the protein structure. By introducing a premature stop-codon, the variant causes the

lack of the last 5 amino acids of the S4D4 segment, the two pore-building segments S5D4 and S6D4 and the entire cytoplasmic C-terminal tail, needed for the protein-protein interaction with one of the β subunits. *SCN3A* has been linked with epilepsy, polymicrogyria, speech and oral motor dysfunction, and ASD has only been indicated as a possible associated feature in *SCN3A*-related NDDs (Brunklaus and Lal 2020). Therefore, until now a direct link between *SCN3A* and ASD has yet to be investigated. An interesting starting point for a deeper exploration of this disease-association could be a functional characterization of the identified *de novo* loss-of-function variant. Indeed, data about *de novo* variants showed in this thesis are preliminary results generated for useful considerations about further investigations that could better delineate the role of established ASD genes as well as highlight the involvement of new ones.

7.3 The contribution of *CACNA1H* biallelic variants in ASD susceptibility

Voltage-gated calcium channels (VGCCs or Cav channels) are transmembrane proteins mediating calcium ions influx into excitable cells upon depolarization of the cell membrane. Specifically, intracellular calcium concentration is mainly regulated by VGCCs activity. They include distinct types of channels, differing in electrophysiological properties. High-voltage-activated (HVA) calcium channels consist of Ca_v1 and Ca_v2 subfamilies, while low-voltage-activated (LVA) calcium channels consist of Ca_v3 subfamily exclusively. The principal functional subunit of Ca_v channel is the pore-forming $\alpha1$ subunit, encoded by the *CACNA1A* to *CACNA1I* and *CACNA1S* genes, and constituting the functional form of LVA or Ca_v3 channels. Differently, in HVA channels (Ca_v1 and Ca_v2) $\alpha1$ subunit requires the presence of $\alpha_2\delta$ and β auxiliary subunits, encoded by the four *CACNA2D1-4* genes and the four *CACNB1-4* genes respectively. The $\alpha1$ subunit co-assembles with one of four $\alpha_2\delta$ and one of four β subunits, forming HVA multiprotein functional complex (Catterall et al. 2005, Zamponi 2016). The main properties of the channel are determined by the pore-forming unit both in LVA and HVA channels, however in the latter group biophysical properties and the pore-forming unit targeting at the cell membrane are profoundly modulated by the auxiliary subunits (Dolphin 2012).

By mediating Ca^{2+} entry, VGCCs are involved in multiple calcium-dependent processes critical for cellular function. Thus, their dysfunction is associated with a wide range of different diseases, including several neuropsychiatric disorders for which VGCCs variants have been indicated as a shared risk factors (Consortium 2013). VGCCs have been

consistently implicated in schizophrenia, ADHD, ASD, epilepsy, bipolar disorder, anxiety and major depressive disorder (MDD), implicating Ca_v channels altered function in alteration of calcium signaling, postsynaptic function, synaptic plasticity and gene transcription (Heyes et al. 2015, Zamponi 2016, Nanou and Catterall 2018, Andrade et al. 2019).

Specifically, both rare and common VGCCs variants have been identified in ASD individuals, and perturbation of intracellular calcium homeostasis caused by disruption of VGCCs genes has been correlated with enhanced ASD susceptibility (Schmunk and Gargus 2013, Reilly et al. 2020, Liao and Li 2020). Moreover, 13 out of 18 VGCCs genes are included in the SFARI Gene database (<https://gene.sfari.org/>), the evolving list of genes implicated in ASD susceptibility. Specifically, 4 VGCCs genes are reported to be clearly implicated in ASD (*CACNA1A*, *CACNA1C*; *CACNA1E* and *CACNA2D3*), while other 3 genes are defined as strong ASD candidate genes (*CACNA1D*, *CACNA1H* and *CACNB2*).

In order to investigate the role of Ca_v channels in ASD aetiology, we looked for rare coding damaging variants in VGCCs genes in our WGS data set. We identified 53 ultra-rare damaging variants in VGCCs genes, none of them was *de novo*. Interestingly, 4 rare damaging biallelic variants were detected in the *CACNA1H* gene in two ASD families.

The *CACNA1H* gene encodes the LVA T-type calcium channel $Ca_v3.2$, that is widely expressed in excitable cells. In brain it is highly expressed in thalamus, hippocampus, amygdala and putamen, where it regulates intracellular calcium concentration, playing important roles in neuronal firing and in neurotransmitter release (Chemin et al. 2000, Perez-Reyes 2003, Carbone, Calorio and Vandael 2014).

CACNA1H is reported to be a strong candidate for ASD as both *de novo* and inherited rare variants in *CACNA1H* were identified in individuals with ASD and other neurodevelopmental disorders, and functional analysis showed a significant effect of *CACNA1H* variants on channel function (Splawski et al. 2006, Iossifov et al. 2014, Iossifov et al. 2015, D'Gama et al. 2015, Takata et al. 2018, Ruzzo et al. 2019, Long et al. 2019, Feliciano et al. 2019, da Silva Montenegro et al. 2020, Weiss and Zamponi 2020). Specifically, Splawski and colleagues tested 5 missense variants identified in a sample of 461 ASD individuals, and detected a decreased activity in mutant channels, implicating *CACNA1H* mutations in ASD phenotype modulation with loss-of-function mechanism and incomplete penetrance (Splawski et al. 2006).

No biallelic variants in *CACNA1H* have been previously reported in individuals with ASD. Compound heterozygous variants affecting *CACNA1H* have been previously identified in neuromuscular disorders, for which functional analysis supported the involvement of loss-of-

function *CACNA1H* variants acting in a recessive inheritance model (Steinberg et al. 2015, Rzhetsky et al. 2016, Carter et al. 2019).

In our WSG cohort the *CACNA1H* gene was affected by biallelic variants in 3 ASD individuals (2 monozygotic twins and 1 independent case) belonging to two families. Probands 22.3-22.4 biallelic variants (Ser2338Phe and Pro2124Leu) affect the cytoplasmic C-terminal domain of the channel, that is not directly implicated in ion transport and channel functionality but that could play a regulatory role by interacting with Syntaxin-1A (Weiss et al. 2012); proband 105.3 variants (Pro849Ser and Lys785Met) are located in critical regions of the protein, thus being expected to have a greater effect on the T-type calcium channel function. Indeed, several variants affecting the same channel region were previously identified and functionally tested, with some of them showing a functional effect (Weiss and Zamponi 2020). Specifically, proband 105.3 Pro849Ser variant is located in the S2 segment of the second protein domain, in which two gain of function mutations were already identified in individuals with Childhood Absence Epilepsy (Khosravani et al. 2004, Peloquin et al. 2006). Differently, the Lys785Met mutation is located in the cytosolic I-II loop, that has been found to be an important regulator of channel function, contributing to the regulation of its gating properties and surface expression (Vitko et al. 2007, Arias-Olguín et al. 2008). Several mutations affecting the I-II loop were identified in individuals affected by idiopathic generalised epilepsy, neuromuscular disorder and chronic pain, with most of them occurring in the region surrounding Lys785Met variant, and some of which showing gain of function effect (Weiss and Zamponi 2020).

Therefore, functional characterization of *CACNA1H* variants was carried out to clarify their effect on the $Ca_v3.2$ channel activity. No loss-of-function effect, suggested by decrease in the voltage-gated Ca^{2+} conductance, was observed for any of the tested variants. By contrast, a clear gain of function effect was observed for the Lys785Met mutant channel, which exhibited a higher functional expression in HEK293 cells, along with a significantly more hyperpolarized V_h value, indicating a higher open probability than WT at the same potential. These properties conferred the ability to transfer, upon activation, a larger electrical charge to the Lys785Met mutant channel when compared to WT, suggesting an increased Ca^{2+} entry in neurons expressing this mutant channel. A detailed kinetic analysis revealed that three out of four mutants (Ser2338Phe, Pro849Ser and Lys785Met) exhibited altered activation kinetics. Hence, in family 22 only one *CACNA1H* variant caused mild alterations in channel properties, while both variants in family 105 showed a functional effect, even if with variable intensity and acting on different channel properties. It is possible that the identified

variants in *CACNA1H* could lead to subtle dysregulation of intracellular Ca^{2+} concentration, thus altering neuronal cell signalling and gene transcription, and contributing to create a high-risk genetic background together with other risk variants. Indeed, several inherited rare damaging variants affecting genes previously implicated in ASD were identified in probands of both families.

More generally, our results further support *CACNA1H* involvement in ASD development, suggesting that *CACNA1H* variants have a relatively small effect on protein function, not sufficient to cause ASD, but they may act additively to drive individual susceptibility over the ASD threshold, thus increasing risk for ASD. This hypothesis is consistent with ASD polygenic inheritance and with previous finding about *CACNA1H* variants in ASD patients, which revealed to modulate phenotypic expression with incomplete penetrance rather than acting as major disease contributors (Splawski et al. 2006).

8. Conclusions

ASD is a heterogeneous and highly heritable neurodevelopmental disorder, with a complex genetic architecture, consisting of a combination of common low-risk and more penetrant rare variants. Specifically, both inherited and *de novo* rare variants highly contribute to individual risk for ASD. The application of NGS technologies has considerably improved rare variants analysis, leading to the identification of more than 100 ASD candidate genes (Iakoucheva et al. 2019, Satterstrom et al. 2020, Mahjani et al. 2021).

This project explored the role of rare variants in ASD susceptibility, taking advantage from NGS strategies. A clinically well-defined cohort of 106 ASD families was collected and genetically characterized by WES or WGS.

In the first part of the project, we explored the contribution of inherited rare variants in the context of a high risk genetic background. Specifically, by WES we characterized a trio family in which a large 3' exonic *NRXNI* deletion is transmitted from the unaffected mother to the ASD child, identifying an increased burden of deleterious variants in the proband that could modulate the CNV penetrance and determine the development of ASD phenotype. In addition, obtained data highlighted the importance of a comprehensive assessment of the genomic landscape of ASD individuals, even in presence of pathogenic variants, as multiple rare variants could contribute in conjunction to the overall genetic risk to the disorder outcome.

In the second part of the project, we investigated the role in ASD aetiology of both *de novo* and inherited rare variants identified through WGS. A robust and efficient variants filtering and prioritization pipeline was developed to efficiently manage and analyse sequencing data. Through the developed pipeline, a stringent set of rare recessive-acting and ultra-rare variants was obtained.

First, it was used to perform a preliminary analysis of rare *de novo* variants, focusing on more deleterious variants, to discover candidate genes for further analyses.

Then, we assessed the role of rare variants in a specific set of known ASD candidate genes, the VGCCs genes. Specifically, we explored the contribution of biallelic variants affecting *CACNA1H*, suggesting that *CACNA1H* mutations may be involved in ASD development by additively combining with other high risk variants.

Therefore, this project shows some of challenges of interpreting the role of rare variants in ASD susceptibility, consistent with previous findings. Ultra-rare more penetrant *de novo* variants in ASD genes could be sufficient alone to determine the phenotypic outcome, while

less penetrant rare and ultra-rare inherited variants could act together with large-effect variants, modulating their penetrance and expressivity, or adding up with hundreds of small-effect variants, to cross the disorder susceptibility threshold on a compromised genetic background.

Further investigations will be necessary to integrate and enhance these findings. Characterization of most promising *de novo* variants may clarify the role of poorly investigated known ASD genes, as well as determine the association with not implicated ones. Enrichment analysis on WGS data from the entire study cohort may provide additional support for the involvement of known ASD pathways and highlight the involvement of new cellular pathways. Moreover, even if their role is more difficult to interpret compared to coding variants, the analysis of non-coding variants may bring to light new types of ASD risk variants. For example, screening of WGS data set for variants affecting cryptic splice sites, identified by machine-learning based predictors, may provide further evidence for this recently proposed new source of ASD risk variations. Additionally, WGS and WES data generated by this project will be integrated with data emerged from ongoing SNP-array analysis, performed on the same cohort of ASD families. This will allow us to obtain a comprehensive characterization of the genomic variation of these families, and to investigate the effect of both sequence and structural rare variants, and their interplay in ASD risk. Finally, the potential contribution of common variants, not explored in this study due to the sample size, will be assessed in collaboration with the University of Washington in the context of a large-scale study, including WGS data from our sample.

Bibliography

- An, J. Y., K. Lin, L. Zhu, D. M. Werling, S. Dong, H. Brand, H. Z. Wang, X. Zhao, G. B. Schwartz, R. L. Collins, B. B. Currall, C. Dastmalchi, J. Dea, C. Duhn, M. C. Gilson, L. Klei, L. Liang, E. Markenscoff-Papadimitriou, S. Pochareddy, N. Ahituv, J. D. Buxbaum, H. Coon, M. J. Daly, Y. S. Kim, G. T. Marth, B. M. Neale, A. R. Quinlan, J. L. Rubenstein, N. Sestan, M. W. State, A. J. Willsey, M. E. Talkowski, B. Devlin, K. Roeder & S. J. Sanders (2018) Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science*, 362.
- Andrade, A., A. Brennecke, S. Mallat, J. Brown, J. Gomez-Rivadeneira, N. Czepiel & L. Londrigan (2019) Genetic Associations between Voltage-Gated Calcium Channels and Psychiatric Disorders. *Int J Mol Sci*, 20.
- Arias-Olguín, I. I., I. Vitko, M. Fortuna, J. P. Baumgart, S. Sokolova, I. A. Shumilin, A. Van Deusen, M. Soriano-García, J. C. Gomora & E. Perez-Reyes (2008) Characterization of the gating brake in the I-II loop of Ca(v)3.2 T-type Ca(2+) channels. *J Biol Chem*, 283, 8136-44.
- Arikkath, J., I. Israely, Y. Tao, L. Mei, X. Liu & L. F. Reichardt (2008) Erbin controls dendritic morphogenesis by regulating localization of delta-catenin. *J Neurosci*, 28, 7047-56.
- Auton, A., L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis & G. P. Consortium (2015) A global reference for human genetic variation. *Nature*, 526, 68-74.
- Avazzadeh, S., K. McDonagh, J. Reilly, Y. Wang, S. D. Boomkamp, V. McInerney, J. Krawczyk, J. Fitzgerald, N. Feerick, M. O'Sullivan, A. Jalali, E. B. Forman, S. A. Lynch, S. Ennis, N. Cosemans, H. Peeters, P. Dockery, T. O'Brien, L. R. Quinlan, L. Gallagher & S. Shen (2019) Increased Ca. *Mol Autism*, 10, 52.
- Bacchelli, E., C. Cameli, M. Viggiano, R. Iglizzi, A. Mancini, R. Tancredi, A. Battaglia & E. Maestrini (2020) An integrated analysis of rare CNV and exome variation in Autism Spectrum Disorder using the Infinium PsychArray. *Sci Rep*, 10, 3198.
- Bai, D., B. H. K. Yip, G. C. Windham, A. Sourander, R. Francis, R. Yoffe, E. Glasson, B. Mahjani, A. Suominen, H. Leonard, M. Gissler, J. D. Buxbaum, K. Wong, D. Schendel, A. Kodesh, M. Breshnahan, S. Z. Levine, E. T. Parner, S. N. Hansen, C. Hultman, A. Reichenberg & S. Sandin (2019) Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. *JAMA Psychiatry*, 76, 1035-1043.
- Battaglia, A., V. Doccini, L. Bernardini, A. Novelli, S. Loddo, A. Capalbo, T. Filippi & J. C. Carey (2013) Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features. *Eur J Paediatr Neurol*, 17, 589-99.

- Bayés, A., L. N. van de Lagemaat, M. O. Collins, M. D. Croning, I. R. Whittle, J. S. Choudhary & S. G. Grant (2011) Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat Neurosci*, 14, 19-21.
- Belyeu, J. R., H. Brand, H. Wang, X. Zhao, B. S. Pedersen, J. Feusier, M. Gupta, T. J. Nicholas, J. Brown, L. Baird, B. Devlin, S. J. Sanders, L. B. Jorde, M. E. Talkowski & A. R. Quinlan (2021) De novo structural mutation rates and gamete-of-origin biases revealed through genome sequencing of 2,396 families. *Am J Hum Genet*, 108, 597-607.
- Bourgeron, T. (2015) From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat Rev Neurosci*, 16, 551-63.
- Brunklaus, A. & D. Lal (2020) Sodium channel epilepsies and neurodevelopmental disorders: from disease mechanisms to clinical application. *Dev Med Child Neurol*, 62, 784-792.
- C Yuen, R. K., D. Merico, M. Bookman, J. L Howe, B. Thiruvahindrapuram, R. V. Patel, J. Whitney, N. Deflaux, J. Bingham, Z. Wang, G. Pellicchia, J. A. Buchanan, S. Walker, C. R. Marshall, M. Uddin, M. Zarrei, E. Deneault, L. D'Abate, A. J. Chan, S. Koyanagi, T. Paton, S. L. Pereira, N. Hoang, W. Engchuan, E. J. Higginbotham, K. Ho, S. Lamoureux, W. Li, J. R. MacDonald, T. Nalpathamkalam, W. W. Sung, F. J. Tsoi, J. Wei, L. Xu, A. M. Tasse, E. Kirby, W. Van Etten, S. Twigger, W. Roberts, I. Drmic, S. Jilderda, B. M. Modi, B. Kellam, M. Szego, C. Cytrynbaum, R. Weksberg, L. Zwaigenbaum, M. Woodbury-Smith, J. Brian, L. Senman, A. Iaboni, K. Doyle-Thomas, A. Thompson, C. Chrysler, J. Leef, T. Savion-Lemieux, I. M. Smith, X. Liu, R. Nicolson, V. Seifer, A. Fedele, E. H. Cook, S. Dager, A. Estes, L. Gallagher, B. A. Malow, J. R. Parr, S. J. Spence, J. Vorstman, B. J. Frey, J. T. Robinson, L. J. Strug, B. A. Fernandez, M. Elsabbagh, M. T. Carter, J. Hallmayer, B. M. Knoppers, E. Anagnostou, P. Szatmari, R. H. Ring, D. Glazer, M. T. Pletcher & S. W. Scherer (2017) Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nat Neurosci*, 20, 602-611.
- Cameli, C., M. Viggiano, M. J. Rochat, A. Maresca, L. Caporali, C. Fiorini, F. Palombo, P. Magini, R. C. Duardo, F. Ceroni, M. C. Scaduto, A. Posar, M. Seri, V. Carelli, P. Visconti, E. Bacchelli & E. Maestrini (2021) An increased burden of rare exonic variants in NRXN1 microdeletion carriers is likely to enhance the penetrance for autism spectrum disorder. *J Cell Mol Med*, 25, 2459-2470.
- Carbone, E., C. Calorio & D. H. Vandael (2014) T-type channel-mediated neurotransmitter release. *Pflugers Arch*, 466, 677-87.
- Carter, M. T., H. J. McMillan, A. Tomin & N. Weiss (2019) Compound heterozygous CACNA1H mutations associated with severe congenital amyotrophy. *Channels (Austin)*, 13, 153-161.
- Catterall, W. A. (2000) From ionic currents to molecular mechanisms: the structure and function of voltage-gated sodium channels. *Neuron*, 26, 13-25.
- Catterall, W. A., E. Perez-Reyes, T. P. Snutch & J. Striessnig (2005) International Union of Pharmacology. XLVIII. Nomenclature and structure-function relationships of voltage-gated calcium channels. *Pharmacol Rev*, 57, 411-25.

- Chemin, J., A. Monteil, C. Briquaire, S. Richard, E. Perez-Reyes, J. Nargeot & P. Lory (2000) Overexpression of T-type calcium channels in HEK-293 cells increases intracellular calcium without affecting cellular proliferation. *FEBS Lett*, 478, 166-72.
- Chiara, M., S. Gioiosa, G. Chillemi, M. D'Antonio, T. Flati, E. Picardi, F. Zambelli, D. S. Horner, G. Pesole & T. Castrignanò (2018) CoVaCS: a consensus variant calling system. *BMC Genomics*, 19, 120.
- Coban-Akdemir, Z., J. J. White, X. Song, S. N. Jhangiani, J. M. Fatih, T. Gambin, Y. Bayram, I. K. Chinn, E. Karaca, J. Punetha, C. Poli, E. Boerwinkle, C. A. Shaw, J. S. Orange, R. A. Gibbs, T. Lappalainen, J. R. Lupski, C. M. B. Carvalho & B.-H. C. f. M. Genomics (2018) Identifying Genes Whose Mutant Transcripts Cause Dominant Disease Traits by Potential Gain-of-Function Alleles. *Am J Hum Genet*, 103, 171-187.
- Consortium, C.-D. G. o. t. P. G. (2013) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet*, 381, 1371-1379.
- Cosemans, N., L. Vandenhove, A. Vogels, K. Devriendt, H. Van Esch, G. Van Buggenhout, H. Olivie, T. de Ravel, E. Ortibus, E. Legius, P. Aerssens, J. Breckpot, J. R. Vermeesch, S. Shen, J. Fitzgerald, L. Gallagher & H. Peeters (2020) The clinical relevance of intragenic. *J Med Genet*, 57, 347-355.
- Cribbs, L. L., J. H. Lee, J. Yang, J. Satin, Y. Zhang, A. Daud, J. Barclay, M. P. Williamson, M. Fox, M. Rees & E. Perez-Reyes (1998) Cloning and characterization of alpha1H from human heart, a member of the T-type Ca²⁺ channel gene family. *Circ Res*, 83, 103-9.
- D'Gama, A. M., S. Pochareddy, M. Li, S. S. Jamuar, R. E. Reiff, A. N. Lam, N. Sestan & C. A. Walsh (2015) Targeted DNA Sequencing from Autism Spectrum Disorder Brains Implicates Multiple Genetic Mechanisms. *Neuron*, 88, 910-917.
- da Silva Montenegro, E. M., C. S. Costa, G. Campos, M. Scliar, T. F. de Almeida, E. C. Zachi, I. M. W. Silva, A. J. S. Chan, M. Zarrei, N. C. V. Lourenço, G. L. Yamamoto, S. Scherer & M. R. Passos-Bueno (2020) Meta-Analyses Support Previous and Novel Autism Candidate Genes: Outcomes of an Unexplored Brazilian Cohort. *Autism Res*, 13, 199-206.
- Darnell, R. B. (2020) The Genetic Control of Stoichiometry Underlying Autism. *Annu Rev Neurosci*, 43, 509-533.
- De Rubeis, S., X. He, A. P. Goldberg, C. S. Poultney, K. Samocha, A. E. Cicek, Y. Kou, L. Liu, M. Fromer, S. Walker, T. Singh, L. Klei, J. Kosmicki, F. Shih-Chen, B. Aleksic, M. Biscaldi, P. F. Bolton, J. M. Brownfeld, J. Cai, N. G. Campbell, A. Carracedo, M. H. Chahrour, A. G. Chiochetti, H. Coon, E. L. Crawford, S. R. Curran, G. Dawson, E. Duketis, B. A. Fernandez, L. Gallagher, E. Geller, S. J. Guter, R. S. Hill, J. Ionita-Laza, P. Jimenez Gonzalez, H. Kilpinen, S. M. Klauck, A. Kolevzon, I. Lee, I. Lei, J. Lei, T. Lehtimäki, C. F. Lin, A. Ma'ayan, C. R. Marshall, A. L. McInnes, B. Neale, M. J. Owen, N. Ozaki, M. Parellada, J. R. Parr, S. Purcell, K. Puura, D. Rajagopalan, K. Rehnström, A. Reichenberg, A. Sabo, M. Sachse, S. J. Sanders, C. Schafer, M. Schulte-Rüther, D. Skuse, C. Stevens, P. Szatmari, K. Tammimies, O. Valladares, A. Voran, W. Li-San, L. A. Weiss, A. J. Willsey, T. W. Yu, R. K. Yuen, E. H. Cook, C. M. Freitag, M. Gill, C. M. Hultman, T. Lehner, A. Palotie, G. D. Schellenberg, P.

- Sklar, M. W. State, J. S. Sutcliffe, C. A. Walsh, S. W. Scherer, M. E. Zwick, J. C. Barrett, D. J. Cutler, K. Roeder, B. Devlin, M. J. Daly, J. D. Buxbaum, D. Study, H. M. C. f. Autism & U. K. Consortium (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature*, 515, 209-15.
- Devlin, B. & S. W. Scherer (2012) Genetic architecture in autism spectrum disorder. *Curr Opin Genet Dev*, 22, 229-37.
- Dolphin, A. C. (2012) Calcium channel auxiliary $\alpha 2\delta$ and β subunits: trafficking and one step beyond. *Nat Rev Neurosci*, 13, 542-55.
- Elliott, A. M., L. R. Simard, G. Coghlan, A. E. Chudley, B. N. Chodirker, C. R. Greenberg, T. Burch, V. Ly, G. M. Hatch & T. Zelinski (2013) A novel mutation in KIAA0196: identification of a gene involved in Ritscher-Schinzel/3C syndrome in a First Nations cohort. *J Med Genet*, 50, 819-22.
- Feliciano, P., X. Zhou, I. Astrovskaya, T. N. Turner, T. Wang, L. Brueggeman, R. Barnard, A. Hsieh, L. G. Snyder, D. M. Muzny, A. Sabo, R. A. Gibbs, E. E. Eichler, B. J. O'Roak, J. J. Michaelson, N. Volfovsky, Y. Shen, W. K. Chung & S. Consortium (2019) Exome sequencing of 457 autism families recruited online provides evidence for autism risk genes. *NPJ Genom Med*, 4, 19.
- Ferri, S. L., T. Abel & E. S. Brodtkin (2018) Sex Differences in Autism Spectrum Disorder: a Review. *Curr Psychiatry Rep*, 20, 9.
- Flaherty, E., S. Zhu, N. Barretto, E. Cheng, P. J. M. Deans, M. B. Fernando, N. Schrode, N. Francoeur, A. Antoine, K. Alganem, M. Halpern, G. Deikus, H. Shah, M. Fitzgerald, I. Ladrán, P. Gochman, J. Rapoport, N. M. Tsankova, R. McCullumsmith, G. E. Hoffman, R. Sebra, G. Fang & K. J. Brennand (2019) Neuronal impact of patient-specific aberrant NRXN1 α splicing. *Nat Genet*, 51, 1679-1690.
- Ginanneschi, F., A. D'Amore, M. Barghigiani, A. Tessa, A. Rossi & F. M. Santorelli (2020) SPG8 mutations in Italian families: clinical data and literature review. *Neurol Sci*, 41, 699-703.
- Girirajan, S., J. A. Rosenfeld, B. P. Coe, S. Parikh, N. Friedman, A. Goldstein, R. A. Filipink, J. S. McConnell, B. Angle, W. S. Meschino, M. M. Nezarati, A. Asamoah, K. E. Jackson, G. C. Gowans, J. A. Martin, E. P. Carmany, D. W. Stockton, R. E. Schnur, L. S. Penney, D. M. Martin, S. Raskin, K. Leppig, H. Thiese, R. Smith, E. Aberg, D. M. Niyazov, L. F. Escobar, D. El-Khechen, K. D. Johnson, R. R. Lebel, K. Siefkas, S. Ball, N. Shur, M. McGuire, C. K. Brasington, J. E. Spence, L. S. Martin, C. Clericuzio, B. C. Ballif, L. G. Shaffer & E. E. Eichler (2012) Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *N Engl J Med*, 367, 1321-31.
- Griesi-Oliveira, K. & A. L. Sertié (2017) Autism spectrum disorders: an updated guide for genetic counseling. *Einstein (Sao Paulo)*, 15, 233-238.
- Grove, J., S. Ripke, T. D. Als, M. Mattheisen, R. K. Walters, H. Won, J. Pallesen, E. Agerbo, O. A. Andreassen, R. Anney, S. Awashti, R. Belliveau, F. Bettella, J. D. Buxbaum, J. Bybjerg-Grauholm, M. Bækvad-Hansen, F. Cerrato, K. Chambert, J. H. Christensen, C. Churchhouse, K. Dellenvall, D. Demontis, S. De Rubeis, B. Devlin, S. Djurovic, A.

- L. Dumont, J. I. Goldstein, C. S. Hansen, M. E. Hauberg, M. V. Hollegaard, S. Hope, D. P. Howrigan, H. Huang, C. M. Hultman, L. Klei, J. Maller, J. Martin, A. R. Martin, J. L. Moran, M. Nyegaard, T. Nærland, D. S. Palmer, A. Palotie, C. B. Pedersen, M. G. Pedersen, T. dPoterba, J. B. Poulsen, B. S. Pourcain, P. Qvist, K. Rehnström, A. Reichenberg, J. Reichert, E. B. Robinson, K. Roeder, P. Roussos, E. Saemundsen, S. Sandin, F. K. Satterstrom, G. Davey Smith, H. Stefansson, S. Steinberg, C. R. Stevens, P. F. Sullivan, P. Turley, G. B. Walters, X. Xu, K. Stefansson, D. H. Geschwind, M. Nordentoft, D. M. Hougaard, T. Werge, O. Mors, P. B. Mortensen, B. M. Neale, M. J. Daly, A. D. Børglum, A. S. D. W. G. o. t. P. G. Consortium, BUPGEN, M. D. D. W. G. o. t. P. G. Consortium & a. R. Team (2019) Identification of common genetic risk variants for autism spectrum disorder. *Nat Genet*, 51, 431-444.
- Guo, H., E. Bettella, P. C. Marcogliese, R. Zhao, J. C. Andrews, T. J. Nowakowski, M. A. Gillentine, K. Hoekzema, T. Wang, H. Wu, S. Jangam, C. Liu, H. Ni, M. H. Willemsen, B. W. van Bon, T. Rinne, S. J. C. Stevens, T. Kleefstra, H. G. Brunner, H. G. Yntema, M. Long, W. Zhao, Z. Hu, C. Colson, N. Richard, C. E. Schwartz, C. Romano, L. Castiglia, M. Bottitta, S. U. Dhar, D. J. Erwin, L. Emrick, B. Keren, A. Afenjar, B. Zhu, B. Bai, P. Stankiewicz, K. Herman, S. Mercimek-Andrews, J. Juusola, A. B. Wilfert, R. Abou Jamra, B. Büttner, H. C. Mefford, A. M. Muir, I. E. Scheffer, B. M. Regan, S. Malone, J. Gecz, J. Cobben, M. M. Weiss, Q. Waisfisz, E. K. Bijlsma, M. J. V. Hoffer, C. A. L. Ruivenkamp, S. Sartori, F. Xia, J. A. Rosenfeld, R. A. Bernier, M. F. Wangler, S. Yamamoto, K. Xia, A. P. A. Stegmann, H. J. Bellen, A. Murgia, E. E. Eichler & U. o. W. C. f. M. Genomics (2019) Disruptive mutations in TANC2 define a neurodevelopmental syndrome associated with psychiatric disorders. *Nat Commun*, 10, 4679.
- Han, S., J. Nam, Y. Li, S. Kim, S. H. Cho, Y. S. Cho, S. Y. Choi, J. Choi, K. Han, Y. Kim, M. Na, H. Kim, Y. C. Bae & E. Kim (2010) Regulation of dendritic spines, spatial memory, and embryonic development by the TANC family of PSD-95-interacting proteins. *J Neurosci*, 30, 15102-12.
- Heyes, S., W. S. Pratt, E. Rees, S. Dahimene, L. Ferron, M. J. Owen & A. C. Dolphin (2015) Genetic disruption of voltage-gated calcium channels in psychiatric and neurological disorders. *Prog Neurobiol*, 134, 36-54.
- Hiatt, S. M., M. L. Thompson, J. W. Prokop, J. M. J. Lawlor, D. E. Gray, E. M. Bebin, T. Rinne, M. Kempers, R. Pfundt, B. W. van Bon, C. Mignot, C. Nava, C. Depienne, L. Kalsner, A. Rauch, P. Joset, R. Bachmann-Gagescu, I. M. Wentzensen, K. McWalter & G. M. Cooper (2019) Deleterious Variation in BRSK2 Associates with a Neurodevelopmental Disorder. *Am J Hum Genet*, 104, 701-708.
- Hoang, N., C. Cytrynbaum & S. W. Scherer (2018) Communicating complex genomic information: A counselling approach derived from research experience with Autism Spectrum Disorder. *Patient Educ Couns*, 101, 352-361.
- Huang, Y. Z., S. Won, D. W. Ali, Q. Wang, M. Tanowitz, Q. S. Du, K. A. Pelkey, D. J. Yang, W. C. Xiong, M. W. Salter & L. Mei (2000) Regulation of neuregulin signaling by PSD-95 interacting with ErbB4 at CNS synapses. *Neuron*, 26, 443-55.
- Hurley, R. S., M. Losh, M. Parlier, J. S. Reznick & J. Piven (2007) The broad autism phenotype questionnaire. *J Autism Dev Disord*, 37, 1679-90.

- Iakoucheva, L. M., A. R. Muotri & J. Sebat (2019) Getting to the Cores of Autism. *Cell*, 178, 1287-1298.
- Iossifov, I., D. Levy, J. Allen, K. Ye, M. Ronemus, Y. H. Lee, B. Yamrom & M. Wigler (2015) Low load for disruptive mutations in autism genes and their biased transmission. *Proc Natl Acad Sci U S A*, 112, E5600-7.
- Iossifov, I., B. J. O'Roak, S. J. Sanders, M. Ronemus, N. Krumm, D. Levy, H. A. Stessman, K. T. Witherspoon, L. Vives, K. E. Patterson, J. D. Smith, B. Paepier, D. A. Nickerson, J. Dea, S. Dong, L. E. Gonzalez, J. D. Mandell, S. M. Mane, M. T. Murtha, C. A. Sullivan, M. F. Walker, Z. Waqar, L. Wei, A. J. Willsey, B. Yamrom, Y. H. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M. C. Schatz, K. Ye, W. R. McCombie, J. Shendure, E. E. Eichler, M. W. State & M. Wigler (2014) The contribution of de novo coding mutations to autism spectrum disorder. *Nature*, 515, 216-21.
- Iossifov, I., M. Ronemus, D. Levy, Z. Wang, I. Hakker, J. Rosenbaum, B. Yamrom, Y. H. Lee, G. Narzisi, A. Leotta, J. Kendall, E. Grabowska, B. Ma, S. Marks, L. Rodgers, A. Stepansky, J. Troge, P. Andrews, M. Bekritsky, K. Pradhan, E. Ghiban, M. Kramer, J. Parla, R. Demeter, L. L. Fulton, R. S. Fulton, V. J. Magrini, K. Ye, J. C. Darnell, R. B. Darnell, E. R. Mardis, R. K. Wilson, M. C. Schatz, W. R. McCombie & M. Wigler (2012) De novo gene disruptions in children on the autistic spectrum. *Neuron*, 74, 285-99.
- Jacquemont, S., B. P. Coe, M. Hersch, M. H. Duyzend, N. Krumm, S. Bergmann, J. S. Beckmann, J. A. Rosenfeld & E. E. Eichler (2014) A higher mutational burden in females supports a "female protective model" in neurodevelopmental disorders. *Am J Hum Genet*, 94, 415-25.
- Jansen, A., G. C. Dieleman, A. B. Smit, M. Verhage, F. C. Verhulst, T. J. C. Polderman & D. Posthuma (2017) Gene-set analysis shows association between FMRP targets and autism spectrum disorder. *Eur J Hum Genet*, 25, 863-868.
- Kang, H. J., Y. I. Kawasawa, F. Cheng, Y. Zhu, X. Xu, M. Li, A. M. Sousa, M. Pletikos, K. A. Meyer, G. Sedmak, T. Guennel, Y. Shin, M. B. Johnson, Z. Krsnik, S. Mayer, S. Fertuzinhos, S. Umlauf, S. N. Ligo, A. Vortmeyer, D. R. Weinberger, S. Mane, T. M. Hyde, A. Huttner, M. Reimers, J. E. Kleinman & N. Sestan (2011) Spatio-temporal transcriptome of the human brain. *Nature*, 478, 483-9.
- Karczewski, K. J., L. C. Francioli, G. Tiao, B. B. Cummings, J. Alföldi, Q. Wang, R. L. Collins, K. M. Laricchia, A. Ganna, D. P. Birnbaum, L. D. Gauthier, H. Brand, M. Solomonson, N. A. Watts, D. Rhodes, M. Singer-Berk, E. M. England, E. G. Seaby, J. A. Kosmicki, R. K. Walters, K. Tashman, Y. Farjoun, E. Banks, T. Poterba, A. Wang, C. Seed, N. Whiffin, J. X. Chong, K. E. Samocha, E. Pierce-Hoffman, Z. Zappala, A. H. O'Donnell-Luria, E. V. Minikel, B. Weisburd, M. Lek, J. S. Ware, C. Vittal, I. M. Armean, L. Bergelson, K. Cibulskis, K. M. Connolly, M. Covarrubias, S. Donnelly, S. Ferriera, S. Gabriel, J. Gentry, N. Gupta, T. Jeandet, D. Kaplan, C. Llanwarne, R. Munshi, S. Novod, N. Petrillo, D. Roazen, V. Ruano-Rubio, A. Saltzman, M. Schleicher, J. Soto, K. Tibbetts, C. Tolonen, G. Wade, M. E. Talkowski, B. M. Neale, M. J. Daly, D. G. MacArthur & G. A. D. Consortium (2020) The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581, 434-443.

- Khosravani, H., C. Altier, B. Simms, K. S. Hamming, T. P. Snutch, J. Mezeyova, J. E. McRory & G. W. Zamponi (2004) Gating effects of mutations in the Cav3.2 T-type calcium channel associated with childhood absence epilepsy. *J Biol Chem*, 279, 9681-4.
- Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper & J. Shendure (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*, 46, 310-5.
- Klei, L., L. L. McClain, B. Mahjani, K. Panayidou, S. De Rubeis, A. S. Grahnat, G. Karlsson, Y. Lu, N. Melhem, X. Xu, A. Reichenberg, S. Sandin, C. M. Hultman, J. D. Buxbaum, K. Roeder & B. Devlin (2021) How rare and common risk variation jointly affect liability for autism spectrum disorder. *Mol Autism*, 12, 66.
- Kong, A., M. L. Frigge, G. Masson, S. Besenbacher, P. Sulem, G. Magnusson, S. A. Gudjonsson, A. Sigurdsson, A. Jonasdottir, W. S. Wong, G. Sigurdsson, G. B. Walters, S. Steinberg, H. Helgason, G. Thorleifsson, D. F. Gudbjartsson, A. Helgason, O. T. Magnusson, U. Thorsteinsdottir & K. Stefansson (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature*, 488, 471-5.
- Krumm, N., B. J. O'Roak, J. Shendure & E. E. Eichler (2014) A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci*, 37, 95-105.
- Krumm, N., T. N. Turner, C. Baker, L. Vives, K. Mohajeri, K. Witherspoon, A. Raja, B. P. Coe, H. A. Stessman, Z. X. He, S. M. Leal, R. Bernier & E. E. Eichler (2015) Excess of rare, inherited truncating mutations in autism. *Nat Genet*, 47, 582-8.
- Lai, M. C., M. V. Lombardo & S. Baron-Cohen (2014) Autism. *Lancet*, 383, 896-910.
- Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H. H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur & E. A. Consortium (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536, 285-91.
- Li, J., M. Shi, Z. Ma, S. Zhao, G. Euskirchen, J. Ziskin, A. Urban, J. Hallmayer & M. Snyder (2014) Integrated systems analysis reveals a molecular network underlying autism spectrum disorders. *Mol Syst Biol*, 10, 774.
- Liao, X. & Y. Li (2020) Genetic associations between voltage-gated calcium channels and autism spectrum disorder: a systematic review. *Mol Brain*, 13, 96.

- Lim, E. T., S. Raychaudhuri, S. J. Sanders, C. Stevens, A. Sabo, D. G. MacArthur, B. M. Neale, A. Kirby, D. M. Ruderfer, M. Fromer, M. Lek, L. Liu, J. Flannick, S. Ripke, U. Nagaswamy, D. Muzny, J. G. Reid, A. Hawes, I. Newsham, Y. Wu, L. Lewis, H. Dinh, S. Gross, L. S. Wang, C. F. Lin, O. Valladares, S. B. Gabriel, M. dePristo, D. M. Altshuler, S. M. Purcell, M. W. State, E. Boerwinkle, J. D. Buxbaum, E. H. Cook, R. A. Gibbs, G. D. Schellenberg, J. S. Sutcliffe, B. Devlin, K. Roeder, M. J. Daly & N. E. S. Project (2013) Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. *Neuron*, 77, 235-42.
- Liu, Y., Y. Liang, A. E. Cicek, Z. Li, J. Li, R. A. Muhle, M. Krenzer, Y. Mei, Y. Wang, N. Knoblauch, J. Morrison, S. Zhao, Y. Jiang, E. Geller, I. Ionita-Laza, J. Wu, K. Xia, J. P. Noonan, Z. S. Sun & X. He (2018) A Statistical Framework for Mapping Risk Genes from De Novo Mutations in Whole-Genome-Sequencing Studies. *Am J Hum Genet*, 102, 1031-1047.
- Livak, K. J. & T. D. Schmittgen (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25, 402-8.
- Long, S., H. Zhou, S. Li, T. Wang, Y. Ma, C. Li, Y. Zhou, S. Zhou, B. Wu & Y. Wang (2019) The Clinical and Genetic Features of Co-occurring Epilepsy and Autism Spectrum Disorder in Chinese Children. *Front Neurol*, 10, 505.
- Lu, Y., Y. Liang, S. Ning, G. Deng, Y. Xie, J. Song, N. Zuo, C. Feng & Y. Qin (2020) Rare partial trisomy and tetrasomy of 15q11-q13 associated with developmental delay and autism spectrum disorder. *Mol Cytogenet*, 13, 21.
- Lyall, K., L. Croen, J. Daniels, M. D. Fallin, C. Ladd-Acosta, B. K. Lee, B. Y. Park, N. W. Snyder, D. Schendel, H. Volk, G. C. Windham & C. Newschaffer (2017) The Changing Epidemiology of Autism Spectrum Disorders. *Annu Rev Public Health*, 38, 81-102.
- Maenner, M. J., K. A. Shaw, J. Baio, A. Washington, M. Patrick, M. DiRienzo, D. L. Christensen, L. D. Wiggins, S. Pettygrove, J. G. Andrews, M. Lopez, A. Hudson, T. Baroud, Y. Schwenk, T. White, C. R. Rosenberg, L. C. Lee, R. A. Harrington, M. Huston, A. Hewitt, A. Esler, J. Hall-Lande, J. N. Poynter, L. Hallas-Muchow, J. N. Constantino, R. T. Fitzgerald, W. Zahorodny, J. Shenouda, J. L. Daniels, Z. Warren, A. Vehorn, A. Salinas, M. S. Durkin, P. M. Dietz, EdS1 & PhD-7 (2020) Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2016. *MMWR Surveill Summ*, 69, 1-12.
- Mahjani, B., S. De Rubeis, C. Gustavsson Mahjani, M. Mulhern, X. Xu, L. Klei, F. K. Satterstrom, J. Fu, M. E. Talkowski, A. Reichenberg, S. Sandin, C. M. Hultman, D. E. Grice, K. Roeder, B. Devlin & J. D. Buxbaum (2021) Prevalence and phenotypic impact of rare potentially damaging variants in autism spectrum disorder. *Mol Autism*, 12, 65.
- Masini, E., E. Loi, A. F. Vega-Benedetti, M. Carta, G. Doneddu, R. Fadda & P. Zavattari (2020) An Overview of the Main Genetic, Epigenetic and Environmental Factors Involved in Autism Spectrum Disorder Focusing on Synaptic Activity. *Int J Mol Sci*, 21.

- Nanou, E. & W. A. Catterall (2018) Calcium Channels, Synaptic Plasticity, and Neuropsychiatric Disease. *Neuron*, 98, 466-481.
- Neale, B. M., Y. Kou, L. Liu, A. Ma'ayan, K. E. Samocha, A. Sabo, C. F. Lin, C. Stevens, L. S. Wang, V. Makarov, P. Polak, S. Yoon, J. Maguire, E. L. Crawford, N. G. Campbell, E. T. Geller, O. Valladares, C. Schafer, H. Liu, T. Zhao, G. Cai, J. Lihm, R. Dannenfelser, O. Jabado, Z. Peralta, U. Nagaswamy, D. Muzny, J. G. Reid, I. Newsham, Y. Wu, L. Lewis, Y. Han, B. F. Voight, E. Lim, E. Rossin, A. Kirby, J. Flannick, M. Fromer, K. Shakir, T. Fennell, K. Garimella, E. Banks, R. Poplin, S. Gabriel, M. DePristo, J. R. Wimbish, B. E. Boone, S. E. Levy, C. Betancur, S. Sunyaev, E. Boerwinkle, J. D. Buxbaum, E. H. Cook, B. Devlin, R. A. Gibbs, K. Roeder, G. D. Schellenberg, J. S. Sutcliffe & M. J. Daly (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, 485, 242-5.
- O'Roak, B. J., L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B. P. Coe, R. Levy, A. Ko, C. Lee, J. D. Smith, E. H. Turner, I. B. Stanaway, B. Vernot, M. Malig, C. Baker, B. Reilly, J. M. Akey, E. Borenstein, M. J. Rieder, D. A. Nickerson, R. Bernier, J. Shendure & E. E. Eichler (2012) Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature*, 485, 246-50.
- Okonechnikov, K., A. Conesa & F. García-Alcalde (2016) Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32, 292-4.
- Omasits, U., C. H. Ahrens, S. Müller & B. Wollscheid (2014) Protter: interactive protein feature visualization and integration with experimental proteomic data. *Bioinformatics*, 30, 884-6.
- Pagnamenta, A. T., E. Bacchelli, M. V. de Jonge, G. Mirza, T. S. Scerri, F. Minopoli, A. Chiocchetti, K. U. Ludwig, P. Hoffmann, S. Paracchini, E. Lowy, D. H. Harold, J. A. Chapman, S. M. Klauck, F. Poustka, R. H. Houben, W. G. Staal, R. A. Ophoff, M. C. O'Donovan, J. Williams, M. M. Nöthen, G. Schulte-Körne, P. Deloukas, J. Ragoussis, A. J. Bailey, E. Maestrini, A. P. Monaco & I. M. G. S. O. A. Consortium (2010) Characterization of a family with rare deletions in CNTNAP5 and DOCK4 suggests novel risk loci for autism and dyslexia. *Biol Psychiatry*, 68, 320-8.
- Peloquin, J. B., H. Khosravani, W. Barr, C. Bladen, R. Evans, J. Mezeyova, D. Parker, T. P. Snutch, J. E. McRory & G. W. Zamponi (2006) Functional analysis of Ca_v3.2 T-type calcium channel mutations linked to childhood absence epilepsy. *Epilepsia*, 47, 655-8.
- Perez-Reyes, E. (2003) Molecular physiology of low-voltage-activated t-type calcium channels. *Physiol Rev*, 83, 117-61.
- Petrovski, S., Q. Wang, E. L. Heinzen, A. S. Allen & D. B. Goldstein (2013) Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet*, 9, e1003709.
- Pinto, D., E. Delaby, D. Merico, M. Barbosa, A. Merikangas, L. Klei, B. Thiruvahindrapuram, X. Xu, R. Ziman, Z. Wang, J. A. Vorstman, A. Thompson, R. Regan, M. Pilorge, G. Pellecchia, A. T. Pagnamenta, B. Oliveira, C. R. Marshall, T. R. Magalhaes, J. K. Lowe, J. L. Howe, A. J. Griswold, J. Gilbert, E. Duketis, B. A. Dombroski, M. V. De Jonge, M. Cuccaro, E. L. Crawford, C. T. Correia, J. Conroy, I.

- C. Conceição, A. G. Chiocchetti, J. P. Casey, G. Cai, C. Cabrol, N. Bolshakova, E. Bacchelli, R. Anney, S. Gallinger, M. Cotterchio, G. Casey, L. Zwaigenbaum, K. Wittemeyer, K. Wing, S. Wallace, H. van Engeland, A. Tryfon, S. Thomson, L. Soorya, B. Rogé, W. Roberts, F. Poustka, S. Mouga, N. Minshew, L. A. McInnes, S. G. McGrew, C. Lord, M. Leboyer, A. S. Le Couteur, A. Kolevzon, P. Jiménez González, S. Jacob, R. Holt, S. Guter, J. Green, A. Green, C. Gillberg, B. A. Fernandez, F. Duque, R. Delorme, G. Dawson, P. Chaste, C. Café, S. Brennan, T. Bourgeron, P. F. Bolton, S. Bölte, R. Bernier, G. Baird, A. J. Bailey, E. Anagnostou, J. Almeida, E. M. Wijsman, V. J. Vieland, A. M. Vicente, G. D. Schellenberg, M. Pericak-Vance, A. D. Paterson, J. R. Parr, G. Oliveira, J. I. Nurnberger, A. P. Monaco, E. Maestrini, S. M. Klauck, H. Hakonarson, J. L. Haines, D. H. Geschwind, C. M. Freitag, S. E. Folstein, S. Ennis, et al. (2014) Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*, 94, 677-94.
- Pinto, D., A. T. Pagnamenta, L. Klei, R. Anney, D. Merico, R. Regan, J. Conroy, T. R. Magalhaes, C. Correia, B. S. Abrahams, J. Almeida, E. Bacchelli, G. D. Bader, A. J. Bailey, G. Baird, A. Battaglia, T. Berney, N. Bolshakova, S. Bölte, P. F. Bolton, T. Bourgeron, S. Brennan, J. Brian, S. E. Bryson, A. R. Carson, G. Casallo, J. Casey, B. H. Chung, L. Cochrane, C. Corsello, E. L. Crawford, A. Crossett, C. Cytrynbaum, G. Dawson, M. de Jonge, R. Delorme, I. Drmic, E. Duketis, F. Duque, A. Estes, P. Farrar, B. A. Fernandez, S. E. Folstein, E. Fombonne, C. M. Freitag, J. Gilbert, C. Gillberg, J. T. Glessner, J. Goldberg, A. Green, J. Green, S. J. Guter, H. Hakonarson, E. A. Heron, M. Hill, R. Holt, J. L. Howe, G. Hughes, V. Hus, R. Iglizzi, C. Kim, S. M. Klauck, A. Kolevzon, O. Korvatska, V. Kustanovich, C. M. Lajonchere, J. A. Lamb, M. Laskawiec, M. Leboyer, A. Le Couteur, B. L. Leventhal, A. C. Lionel, X. Q. Liu, C. Lord, L. Lotspeich, S. C. Lund, E. Maestrini, W. Mahoney, C. Mantoulan, C. R. Marshall, H. McConachie, C. J. McDougle, J. McGrath, W. M. McMahon, A. Merikangas, O. Migita, N. J. Minshew, G. K. Mirza, J. Munson, S. F. Nelson, C. Noakes, A. Noor, G. Nygren, G. Oliveira, K. Papanikolaou, J. R. Parr, B. Parrini, T. Paton, A. Pickles, M. Pilorge, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, 466, 368-72.
- Pizzo, L., M. Jensen, A. Polyak, J. A. Rosenfeld, K. Mannik, A. Krishnan, E. McCready, O. Pichon, C. Le Caignec, A. Van Dijk, K. Pope, E. Voorhoeve, J. Yoon, P. Stankiewicz, S. W. Cheung, D. Pazuchanics, E. Huber, V. Kumar, R. L. Kember, F. Mari, A. Curró, L. Castiglia, O. Galesi, E. Avola, T. Mattina, M. Fichera, L. Mandarà, M. Vincent, M. Nizon, S. Mercier, C. Bénéteau, S. Blesson, D. Martin-Coignard, A. L. Mosca-Boidron, J. H. Caberg, M. Bucan, S. Zeesman, M. J. M. Nowaczyk, M. Lefebvre, L. Faivre, P. Callier, C. Skinner, B. Keren, C. Perrine, P. Prontera, N. Marle, A. Renieri, A. Reymond, R. F. Kooy, B. Isidor, C. Schwartz, C. Romano, E. Siermans, D. J. Amor, J. Andrieux & S. Girirajan (2019) Rare variants in the genetic background modulate cognitive and developmental phenotypes in individuals carrying disease-associated variants. *Genet Med*, 21, 816-825.
- Reilly, J., L. Gallagher, G. Leader & S. Shen (2020) Coupling of autism genes to tissue-wide expression and dysfunction of synapse, calcium signalling and transcriptional regulation. *PLoS One*, 15, e0242773.
- Robinson, E. B., K. E. Samocha, J. A. Kosmicki, L. McGrath, B. M. Neale, R. H. Perlis & M. J. Daly (2014) Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc Natl Acad Sci U S A*, 111, 15161-5.

- Ruzzo, E. K., L. Pérez-Cano, J. Y. Jung, L. K. Wang, D. Kashef-Haghighi, C. Hartl, C. Singh, J. Xu, J. N. Hoekstra, O. Leventhal, V. M. Leppä, M. J. Gandal, K. Paskov, N. Stockham, D. Polioudakis, J. K. Lowe, D. A. Prober, D. H. Geschwind & D. P. Wall (2019) Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell*, 178, 850-866.e26.
- Rzhetsky, Y., J. Lazniewska, I. Blesneac, R. Pamphlett & N. Weiss (2016) CACNA1H missense mutations associated with amyotrophic lateral sclerosis alter Cav3.2 T-type calcium channel activity and reticular thalamic neuron firing. *Channels (Austin)*, 10, 466-77.
- Sanders, S. J., M. T. Murtha, A. R. Gupta, J. D. Murdoch, M. J. Raubeson, A. J. Willsey, A. G. Ercan-Sencicek, N. M. DiLullo, N. N. Parikshak, J. L. Stein, M. F. Walker, G. T. Ober, N. A. Teran, Y. Song, P. El-Fishawy, R. C. Murtha, M. Choi, J. D. Overton, R. D. Bjornson, N. J. Carriero, K. A. Meyer, K. Bilguvar, S. M. Mane, N. Sestan, R. P. Lifton, M. Günel, K. Roeder, D. H. Geschwind, B. Devlin & M. W. State (2012) De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*, 485, 237-41.
- Satterstrom, F. K., J. A. Kosmicki, J. Wang, M. S. Breen, S. De Rubeis, J. Y. An, M. Peng, R. Collins, J. Grove, L. Klei, C. Stevens, J. Reichert, M. S. Mulhern, M. Artomov, S. Gerges, B. Sheppard, X. Xu, A. Bhaduri, U. Norman, H. Brand, G. Schwartz, R. Nguyen, E. E. Guerrero, C. Dias, C. Betancur, E. H. Cook, L. Gallagher, M. Gill, J. S. Sutcliffe, A. Thurm, M. E. Zwick, A. D. Børglum, M. W. State, A. E. Cicek, M. E. Talkowski, D. J. Cutler, B. Devlin, S. J. Sanders, K. Roeder, M. J. Daly, J. D. Buxbaum, A. S. Consortium & i.-B. Consortium (2020) Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*, 180, 568-584.e23.
- Schmunk, G. & J. J. Gargus (2013) Channelopathy pathogenesis in autism spectrum disorders. *Front Genet*, 4, 222.
- Skuse, D. H., W. P. Mandy & J. Scourfield (2005) Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *Br J Psychiatry*, 187, 568-72.
- Splawski, I., D. S. Yoo, S. C. Stotz, A. Cherry, D. E. Clapham & M. T. Keating (2006) CACNA1H mutations in autism spectrum disorders. *J Biol Chem*, 281, 22085-22091.
- Steinberg, K. M., B. Yu, D. C. Koboldt, E. R. Mardis & R. Pamphlett (2015) Exome sequencing of case-unaffected-parents trios reveals recessive and de novo genetic variants in sporadic ALS. *Sci Rep*, 5, 9124.
- Stessman, H. A., B. Xiong, B. P. Coe, T. Wang, K. Hoekzema, M. Fenckova, M. Kvarnung, J. Gerds, S. Trinh, N. Cosemans, L. Vives, J. Lin, T. N. Turner, G. Santen, C. Ruivenkamp, M. Kriek, A. van Haeringen, E. Aten, K. Friend, J. Liebelt, C. Barnett, E. Haan, M. Shaw, J. Gecz, B. M. Anderlid, A. Nordgren, A. Lindstrand, C. Schwartz, R. F. Kooy, G. Vandeweyer, C. Helmsmoortel, C. Romano, A. Alberti, M. Vinci, E. Avola, S. Giusto, E. Courchesne, T. Pramparo, K. Pierce, S. Nalabolu, D. G. Amaral, I. E. Scheffer, M. B. Delatycki, P. J. Lockhart, F. Hormozdiari, B. Harich, A. Castells-Nobau, K. Xia, H. Peeters, M. Nordenskjöld, A. Schenck, R. A. Bernier & E. E.

- Eichler (2017) Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nat Genet*, 49, 515-526.
- Sullivan, P. F. & D. H. Geschwind (2019) Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell*, 177, 162-183.
- Szklarczyk, D., A. L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N. T. Doncheva, J. H. Morris, P. Bork, L. J. Jensen & C. V. Mering (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*, 47, D607-D613.
- Südhof, T. C. (2017) Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. *Cell*, 171, 745-769.
- Takata, A., N. Miyake, Y. Tsurusaki, R. Fukai, S. Miyatake, E. Koshimizu, I. Kushima, T. Okada, M. Morikawa, Y. Uno, K. Ishizuka, K. Nakamura, M. Tsujii, T. Yoshikawa, T. Toyota, N. Okamoto, Y. Hiraki, R. Hashimoto, Y. Yasuda, S. Saitoh, K. Ohashi, Y. Sakai, S. Ohga, T. Hara, M. Kato, A. Ito, C. Seiwa, E. Shirahata, H. Osaka, A. Matsumoto, S. Takeshita, J. Tohyama, T. Saikusa, T. Matsuishi, T. Nakamura, T. Tsuboi, T. Kato, T. Suzuki, H. Saitsu, M. Nakashima, T. Mizuguchi, F. Tanaka, N. Mori, N. Ozaki & N. Matsumoto (2018) Integrative Analyses of De Novo Mutations Provide Deeper Biological Insights into Autism Spectrum Disorder. *Cell Rep*, 22, 734-747.
- Talavera, K., M. Staes, A. Janssens, N. Klugbauer, G. Droogmans, F. Hofmann & B. Nilius (2001) Aspartate residues of the Glu-Glu-Asp-Asp (EEDD) pore locus control selectivity and permeation of the T-type Ca(2+) channel alpha(1G). *J Biol Chem*, 276, 45628-35.
- Tick, B., P. Bolton, F. Happé, M. Rutter & F. Rijdsdijk (2016) Heritability of autism spectrum disorders: a meta-analysis of twin studies. *J Child Psychol Psychiatry*, 57, 585-95.
- Traut, W., D. Weichenhan, H. Himmelbauer & H. Winking (2006) New members of the neurexin superfamily: multiple rodent homologues of the human CASPR5 gene. *Mamm Genome*, 17, 723-31.
- Turner, T. N., B. P. Coe, D. E. Dickel, K. Hoekzema, B. J. Nelson, M. C. Zody, Z. N. Kronenberg, F. Hormozdiari, A. Raja, L. A. Pennacchio, R. B. Darnell & E. E. Eichler (2017) Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell*, 171, 710-722.e12.
- Turner, T. N., F. Hormozdiari, M. H. Duyzend, S. A. McClymont, P. W. Hook, I. Iossifov, A. Raja, C. Baker, K. Hoekzema, H. A. Stessman, M. C. Zody, B. J. Nelson, J. Huddleston, R. Sandstrom, J. D. Smith, D. Hanna, J. M. Swanson, E. M. Faustman, M. J. Bamshad, J. Stamatoyannopoulos, D. A. Nickerson, A. S. McCallion, R. Darnell & E. E. Eichler (2016) Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. *Am J Hum Genet*, 98, 58-74.
- Vaags, A. K., A. C. Lionel, D. Sato, M. Goodenberger, Q. P. Stein, S. Curran, C. Ogilvie, J. W. Ahn, I. Drmic, L. Senman, C. Chrysler, A. Thompson, C. Russell, A. Prasad, S. Walker, D. Pinto, C. R. Marshall, D. J. Stavropoulos, L. Zwaigenbaum, B. A.

- Fernandez, E. Fombonne, P. F. Bolton, D. A. Collier, J. C. Hodge, W. Roberts, P. Szatmari & S. W. Scherer (2012) Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am J Hum Genet*, 90, 133-41.
- Vitko, I., I. Bidaud, J. M. Arias, A. Mezghrani, P. Lory & E. Perez-Reyes (2007) The I-II loop controls plasma membrane expression and gating of Ca(v)3.2 T-type Ca²⁺ channels: a paradigm for childhood absence epilepsy mutations. *J Neurosci*, 27, 322-30.
- Vorstman, J. A. S., J. R. Parr, D. Moreno-De-Luca, R. J. L. Anney, J. I. Nurnberger & J. F. Hallmayer (2017) Autism genetics: opportunities and challenges for clinical translation. *Nat Rev Genet*, 18, 362-376.
- Wang, K., M. Li & H. Hakonarson (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 38, e164.
- Weiner, D. J., E. M. Wigdor, S. Ripke, R. K. Walters, J. A. Kosmicki, J. Grove, K. E. Samocha, J. I. Goldstein, A. Okbay, J. Bybjerg-Grauholm, T. Werge, D. M. Hougaard, J. Taylor, D. Skuse, B. Devlin, R. Anney, S. J. Sanders, S. Bishop, P. B. Mortensen, A. D. Børglum, G. D. Smith, M. J. Daly, E. B. Robinson, i.-B. A. Group & P. G. C. A. Group (2017) Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet*, 49, 978-985.
- Weiss, N., S. Hameed, J. M. Fernández-Fernández, K. Fablet, M. Karmazinova, C. Poillot, J. Proft, L. Chen, I. Bidaud, A. Monteil, S. Huc-Brandt, L. Lacinova, P. Lory, G. W. Zamponi & M. De Waard (2012) A Ca(v)3.2/syntaxin-1A signaling complex controls T-type channel activity and low-threshold exocytosis. *J Biol Chem*, 287, 2810-8.
- Weiss, N. & G. W. Zamponi (2020) Genetic T-type calcium channelopathies. *J Med Genet*, 57, 1-10.
- Wen, Y., M. J. Alshikho & M. R. Herbert (2016) Pathway Network Analyses for Autism Reveal Multisystem Involvement, Major Overlaps with Other Diseases and Convergence upon MAPK and Calcium Signaling. *PLoS One*, 11, e0153329.
- Wiggins, L. D., A. Reynolds, C. E. Rice, E. J. Moody, P. Bernal, L. Blaskey, S. A. Rosenberg, L. C. Lee & S. E. Levy (2015) Using standardized diagnostic instruments to classify children with autism in the study to explore early development. *J Autism Dev Disord*, 45, 1271-80.
- Wilfert, A. B., T. N. Turner, S. C. Murali, P. Hsieh, A. Sulovari, T. Wang, B. P. Coe, H. Guo, K. Hoekzema, T. E. Bakken, L. H. Winterkorn, U. S. Evani, M. Byrska-Bishop, R. K. Earl, R. A. Bernier, M. C. Zody, E. E. Eichler & S. Consortium (2021) Recent ultra-rare inherited variants implicate new autism candidate risk genes. *Nat Genet*, 53, 1125-1134.
- Wu, H., H. Li, T. Bai, L. Han, J. Ou, G. Xun, Y. Zhang, Y. Wang, G. Duan, N. Zhao, B. Chen, X. Du, M. Yao, X. Zou, J. Zhao, Z. Hu, E. E. Eichler, H. Guo & K. Xia (2020) Phenotype-to-genotype approach reveals head-circumference-associated genes in an autism spectrum disorder cohort. *Clin Genet*, 97, 338-346.

- Yu, T. W., M. H. Chahrour, M. E. Coulter, S. Jiralerspong, K. Okamura-Ikeda, B. Ataman, K. Schmitz-Abe, D. A. Harmin, M. Adli, A. N. Malik, A. M. D'Gama, E. T. Lim, S. J. Sanders, G. H. Mochida, J. N. Partlow, C. M. Sunu, J. M. Felie, J. Rodriguez, R. H. Nasir, J. Ware, R. M. Joseph, R. S. Hill, B. Y. Kwan, M. Al-Saffar, N. M. Mukaddes, A. Hashmi, S. Balkhy, G. G. Gascon, F. M. Hisama, E. LeClair, A. Poduri, O. Oner, S. Al-Saad, S. A. Al-Awadi, L. Bastaki, T. Ben-Omran, A. S. Teebi, L. Al-Gazali, V. Eapen, C. R. Stevens, L. Rappaport, S. B. Gabriel, K. Markianos, M. W. State, M. E. Greenberg, H. Taniguchi, N. E. Braverman, E. M. Morrow & C. A. Walsh (2013) Using whole-exome sequencing to identify inherited causes of autism. *Neuron*, 77, 259-73.
- Yuen, R. K., B. Thiruvahindrapuram, D. Merico, S. Walker, K. Tammimies, N. Hoang, C. Chrysler, T. Nalpathamkalam, G. Pellecchia, Y. Liu, M. J. Gazzellone, L. D'Abate, E. Deneault, J. L. Howe, R. S. Liu, A. Thompson, M. Zarrei, M. Uddin, C. R. Marshall, R. H. Ring, L. Zwaigenbaum, P. N. Ray, R. Weksberg, M. T. Carter, B. A. Fernandez, W. Roberts, P. Szatmari & S. W. Scherer (2015) Whole-genome sequencing of quartet families with autism spectrum disorder. *Nat Med*, 21, 185-91.
- Zamponi, G. W. (2016) Targeting voltage-gated calcium channels in neurological and psychiatric diseases. *Nat Rev Drug Discov*, 15, 19-34.
- Zarrei, M., J. R. MacDonald, D. Merico & S. W. Scherer (2015) A copy number variation map of the human genome. *Nat Rev Genet*, 16, 172-83.

Appendix

Table S1. Ultra-rare deleterious variants in VGCCs genes identified in our WGS data set. ASD affected individuals are indicated in bold.

Genomic change (hg38)	Amino acid change	Effect (CADD score)	Individual ID (Inheritance)	dbSNP	MAF (gnomAD v.3.0)	Gene (SFARI score)
NC_000002.12:g.151860740A>G	NP_001005747.1:p.(Ile246Thr)	missense (27.4)	98.3 (Paternal)	rs765961368	0.00000698	<i>CACNB4</i>
NC_000009.12:g.138023642C>T	NP_000709.1:p.(Arg967Trp)	missense (18)	66.3 (Maternal)	rs200418906	0.00004250	<i>CACNA1B</i> (3)
NC_000009.12:g.138121659G>A	NP_000709.1:p.(Arg2227Gln)	missense (24.6)	31.3 -31.4 (Paternal)	.	0.00000698	<i>CACNA1B</i> (3)
NC_000009.12:g.138121688G>A	NP_000709.1:p.(Ala2237Thr)	missense (15.73)	87.3 -87.4 (Maternal)	rs77641565	0.00030000	<i>CACNA1B</i> (3)
NC_000009.12:g.138121809C>T	NP_000709.1:p.(Thr2277Ile)	missense (22.9)	60.3 (Paternal)	rs760115118	0.00006980	<i>CACNA1B</i> (3)
NC_000009.12:g.138121907G>A	NP_000709.1:p.(Val2310Met)	missense (22.4)	18.3 (Paternal)	rs201689533	0.00002090	<i>CACNA1B</i> (3)
NC_000023.11:g.49211360C>T	NP_005174.2:p.(Ala1419Thr)	missense (18.34)	5.3 (Maternal)	rs782741094	0.00020000	<i>CACNA1F</i> (3)
NC_000023.11:g.49211983C>T	NP_005174.2:p.(Gly1350Ser)	missense (24.7)	112.3 (Maternal)	rs782780521	0.00002840	<i>CACNA1F</i> (3)
NC_000001.11:g.181790533C>A	NP_001192222.1:p.(Gln1959Lys)	missense (16.89)	43.3 (Paternal)	rs761958333	.	<i>CACNA1E</i> (1)
NC_000001.11:g.201048619_201048620del	NP_000060.2:p.(Phe1468Cysfs*14)	frameshift deletion (.)	40.4 (Maternal)	.	.	<i>CACNA1S</i>
NC_000001.11:g.201051096T>C	NP_000060.2:p.(Tyr1334Cys)	missense (25)	46.3 -46.4 (Maternal)	rs146158332	0.00006980	<i>CACNA1S</i>
NC_000001.11:g.201062475C>G	NP_000060.2:p.(Glu965Gln)	missense (21)	49.3 (Maternal)	rs762350071	0.00004890	<i>CACNA1S</i>
NC_000001.11:g.201075539A>G	NP_000060.2:p.(Met635Thr)	missense (23.2)	14.4 (Paternal)	rs144590408	0.00002800	<i>CACNA1S</i>
NC_000003.12:g.50366093C>T	NP_001005505.1:p.(Arg927His)	missense (23.3)	101.3 (Maternal)	rs761523126	0.00003490	<i>CACNA2D2</i>
NC_000003.12:g.53775936G>A	NP_000711.1:p.(Gly1438Glu)	missense (26.4)	54.3 (Paternal)	.	.	<i>CACNA1D</i> (2)
NC_000003.12:g.54627841A>G	NP_060868.2:p.(Ile340Val)	missense (19.62)	54.3 -54.4 (Maternal)	rs185055678	0.00009770	<i>CACNA2D3</i> (1)
NC_000003.12:g.54764331G>A	NP_060868.2:p.(Glu454Lys)	missense (26.6)	103.3 (Maternal)	rs780345468	0.00001400	<i>CACNA2D3</i> (1)
NC_000003.12:g.55018224C>T	NP_060868.2:p.(Thr965Ile)	missense (22.4)	68.3 - 68.4 -68.5 (Paternal)	.	0.00000698	<i>CACNA2D3</i> (1)
NC_000007.14g.81967601C>G	NP_000713.2:p.(Asp820His)	missense (22.3)	54.3 -54.4 (Maternal)	.	.	<i>CACNA2D1</i> (3)
NC_000010.11g.18514277G>C	NP_000715.2:p.(Aap183His)	missense (28.5)	103.3 (Maternal)	.	.	<i>CACNB2</i> (2)
NC_000010.11g.18514515G>T	NP_963887.2:p.(Lys225Asn)	missense (22.1)	109.3 (Maternal)	rs199539261	0.00030000	<i>CACNB2</i> (2)
NC_000010.11g.18538234C>T	NP_000715.2:p.(Leu398Phe)	missense (24.7)	109.3 (Maternal)	rs145638628	0.00030000	<i>CACNB2</i> (2)
NC_000010.11g.18539558G>C	NP_000715.2:p.(Arg551Pro)	missense (25.2)	56.3 (Paternal)	rs577739840	0.00008410	<i>CACNB2</i> (2)

NC_000012.12:g. 1793753C>T	NP_758952.4: p.(Ala1106Thr)	missense (24.5)	22.3-22.4 (Paternal)	.	.	<i>CACNA2D4</i>
NC_000012.12:g. 1854027G>A	NP_758952.4: p.(Arg724Trp)	missense (32)	12.5 (Paternal)	rs371178386	0.00010000	<i>CACNA2D4</i>
NC_000012.12:g. 1875298T>C	NP_758952.4: p.(Asn587Asp)	missense (28.3)	99.4 (Maternal)	rs759153873	.	<i>CACNA2D4</i>
NC_000012.12:g. 1886224G>A	NP_758952.4: p.(Ala331Val)	missense (32)	22.3-22.4 (Paternal)	.	0.00001400	<i>CACNA2D4</i>
NC_000012.12:g. 1913025G>A	NP_758952.4: p.(Gln142Ter)	stopgain (34)	60.3 (Paternal)	rs776189698	.	<i>CACNA2D4</i>
NC_000012.12:g. 2493262C>T	NP_955630.3: p.(Thr330Met)	missense (23.1)	25.3 (Maternal)	rs377345545	0.00002090	<i>CACNA1C</i> (1)
NC_000012.12:g. 2679514C>G	NP_955630.3: p.(Thr1769Ser)	missense (20.1)	16.3 (Maternal)	rs554737427	0.00002090	<i>CACNA1C</i> (1)
NC_000016.10:g. 1202195T>C	NP_066921.2: p.(Ile582Thr)	missense (19.97)	60.3 (Paternal)	.	0.00000698	<i>CACNA1H</i> (2)
NC_000016.10:g. 1204336C>T	NP_066921.2: p.(Arg777Cys)	missense (19.6)	45.3 (Paternal)	rs375325893	0.00020000	<i>CACNA1H</i> (2)
NC_000016.10:g. 1205117G>A	NP_066921.2: p.(Glu819Lys)	missense (21.2)	34.3 (Maternal)	rs375165169	0.00004190	<i>CACNA1H</i> (2)
NC_000016.10:g. 1205207C>T	NP_066921.2: p.(Pro849Ser)	missense (18.22)	105.3 (Paternal)	rs370675810	0.00009770	<i>CACNA1H</i> (2)
NC_000016.10:g. 1208193A>T	NP_066921.2: p.(Asp1112Val)	missense (18.84)	42.3-42.4 (Maternal)	rs202114960	.	<i>CACNA1H</i> (2)
NC_000016.10:g. 1209251C>T	NP_066921.2: p.(Arg1195Trp)	missense (22.8)	15.3 (Paternal)	rs576035669	0.00006980	<i>CACNA1H</i> (2)
NC_000016.10:g. 1210913G>A	NP_066921.2: p.(Ala1389Thr)	missense (23.1)	62.4 (Maternal)	rs758458100	0.00001400	<i>CACNA1H</i> (2)
NC_000016.10:g. 1215066G>A	NP_066921.2: p.(Arg1675Gln)	missense (31)	17.3 (Maternal)	rs149367557	0.00006280	<i>CACNA1H</i> (2)
NC_000016.10:g. 1220176C>T	NP_066921.2: p.(Arg2082Trp)	missense (18.31)	92.3 (Maternal)	rs771719773	0.00004190	<i>CACNA1H</i> (2)
NC_000016.10:g. 1220652_122065 3insG	NP_066921.2: p.(Asp2243Glyfs* 17)	frameshift insertion (.)	73.3 (Maternal)	rs757244810	0.00004900	<i>CACNA1H</i> (2)
NC_000016.10:g. 1220945C>T	NP_066921.2: p.(Ser2338Phe)	missense (16.99)	22.3-22.4- 22.5 (Paternal)	rs757713867	0.00002790	<i>CACNA1H</i> (2)
NC_000017.11:g. 39175200A>T	NP_000714.3: p.(Ile597Asn)	missense (23.8)	22.3-22.4 (Paternal)	.	.	<i>CACNB1</i> (3)
NC_000017.11:g. 39187570T>C	NP_000714.3: p.(Asn108Ser)	missense (23.4)	28.3-28.4 (Paternal)	rs768351592	0.00001400	<i>CACNB1</i> (3)
NC_000017.11:g. 50569291A>T	NP_061496.2: p.(Ile161Phe)	missense (22.5)	48.3 (Paternal)	rs368561457	.	<i>CACNA1G</i> (3)
NC_000017.11:g. 50575862G>A	NP_061496.2: p.(Arg487His)	missense (22.6)	34.3 (Maternal)	rs373257429	0.00002090	<i>CACNA1G</i> (3)
NC_000017.11:g. 50596835G>A	NP_061496.2: p.(Gly1057Asp)	missense (22.5)	66.3 (Maternal)	.	0.00001400	<i>CACNA1G</i> (3)
NC_000017.11:g. 50626104C>G	NP_061496.2: p.(Leu2163Val)	missense (16.31)	5.3 (Paternal)	.	.	<i>CACNA1G</i> (3)
NC_000019.10:g. 13208986T>C	NP_001120693.1: p.(Thr2185Ala)	missense (22.3)	113.4-113.5 (Unknown)	rs1028538547	0.00002100	<i>CACNA1A</i> (1-S)
NC_000019.10:g. 13214241C>T	NP_001120693.1: p.(Glu1979Lys)	missense (26.8)	94.3 (Maternal)	.	.	<i>CACNA1A</i> (1-S)
NC_000019.10:g. 13214276C>T	NP_001120693.1: p.(Arg1967Gln)	missense (27.2)	4.3 (Maternal)	rs199886234	0.00040000	<i>CACNA1A</i> (1-S)
NC_000022.11:g. 39570837C>T	NP_066919.2: p.(Arg29Trp)	missense (22.6)	16.3 (Maternal)	rs760018532	0.00004890	<i>CACNA1I</i> (3)
NC_000022.11:g. 39672234G>A	NP_066919.2: p.(Met1525Ile)	missense (21.2)	4.3 (Paternal)	rs759135846	0.00000698	<i>CACNA1I</i> (3)

NC_000022.11:g. 39679369G>A	NP_066919.2: p.(Gly1773Asp)	missense (15.89)	14.3 (Maternal)	rs958740018	0.00001400	<i>CACNA1I</i> (3)
--------------------------------	--------------------------------	---------------------	---------------------------	-------------	------------	-----------------------

Table S2. Relevant ultra-rare deleterious variants identified in families 22 and 105. ASD affected individuals are indicated in bold.

Genomic change (hg38)	Amino acid change	Effect (CADD score)	Individual ID (Inheritance)	dbSNP	MAF (gnomAD v.3.0)	Gene (SFARI score)
a) Family 22						
<i>NC_000006.12:g. 138336377T>G</i>	<i>NP_065073.3: p.(Val2142Gly)</i>	<i>missense (27.1)</i>	22.3-22.4 (<i>De novo</i>)	.	.	<i>ARFGEF3</i>
NC_000001.11:g. 10653490G>A	NP_001073312.1: p.(Pro856Leu)	missense (20.1)	22.3-22.4-22.5 (Paternal)	rs367611106	.	<i>CASZ1</i> (1)
NC_000011.10:g. 119341673G>A	NP_113621.1: p.(R539C)	missense (31)	22.3-22.4 (Paternal)	rs374823079	0.00003488	<i>MFRP</i> (2)
NC_000014.9:g.1 02947795C>T	NP_006026.3: p.(E1153K)	missense (24.4)	22.3-22.4 (Paternal)	rs539521711	0.00004888	<i>CDC42BPB</i> (2)
NC_000016.10:g. 1220945C>T	NP_066921.2: p.(Ser2338Phe)	missense (16.99)	22.3-22.4 (Paternal)	rs757713867	0.00002792	<i>CACNA1H</i> (2)
NC_000019.10:g. 17203202C>T	NP_004136.2: p.(S1645Leu)	missense (22.8)	22.3-22.4-22.5 (Maternal)	rs762171057	0.00000697 9	<i>MYO9B</i> (2)
NC_000005.10:g. 153698985G>A	NP_000818.2: p.(R455H)	missense (26.9)	22.3-22.4-22.5 (Paternal)	rs759117420	0.00002094	<i>GRIA1</i> (2)
NC_000009.12:g. 116975233C>T	NP_054729.3: p.(Val571I)	missense (24.8)	22.3-22.4 (Paternal)	rs752549961	0.00002094	<i>ASTN2</i> (2)
NC_000001.11:g. 190098059G>C	NP_950252.1: p.(Pro754A)	missense (23.8)	22.3-22.4-22.5 (Maternal)	rs751325612	.	<i>BRINP3</i> (3)
NC_000012.12:g. 123838461G>A	NP_997320.2: p.(M1518I)	missense (25.1)	22.3-22.4-22.5 (Paternal)	rs750213588	0.00002094	<i>DNAH10</i> (3)
NC_000017.11:g. 10459346C>T	NP_060003.2: p.(Val498M)	missense (25.7)	22.3-22.4 (Paternal)	rs143641715	0.00009777	<i>MYH4</i> (3)
NC_000017.11:g. 39175200A>T	NP_000714.3: p.(Ile597Asn)	missense (23.8)	22.3-22.4 (Paternal)	.	.	<i>CACNB1</i> (3)
NC_000017.11:g. 63945452C>T	NP_000325.4: p.(Glu1210Lys)	missense (19.44)	22.3-22.4 (Paternal)	rs761135349	0.00002094	<i>SCN4A</i> (3)
NC_000020.11:g. 53953667G>A	NP_003648.2: p.(Thr482Ile)	missense (17.26)	22.3-22.4-22.5 (Maternal)	.	.	<i>BCAS1</i> (3)
NC_000003.12:g. 175097095A>G	NP_996898.2: p.(Ser117Gly)	missense (18.11)	22.3-22.4 (Paternal)	rs746178036	.	<i>NAALADL2</i> (3)
NC_000006.12:g. 166459476G>T	NP_066958.2: p.(Pro350Thr)	missense (18.58)	22.3-22.4-22.5 (Paternal)	.	.	<i>RPS6KA2</i> (3)
NC_000007.14:g. 128030912C>T	.	splicing (.)	22.3-22.4-22.5 (Paternal)	.	0.00000698 2	<i>LRRC4</i> (3)
NC_000008.11:g. 143867209insC	NP_112598.3: p.(Val2016Cysfs*82)	frameshift insertion (.)	22.3-22.4 (Maternal)	rs781861594	0.00002817	<i>EPPK1</i> (3)
NC_000006.12:g. 152453654C>T	NP_892006.3: p.(Asp987Asn)	missense (24.4)	22.3-22.4-22.5 (Paternal)	.	0.00001396	<i>SYNE1</i> (3-S)
NC_000013.11:g. 100209381C>G	NP_000273.2: p.(Ala173Gly)	missense (23)	22.3-22.4 (Paternal)	rs750928504	0.00002094	<i>PCCA</i> (S)
NC_000015.10:g. 28142369G>C	NP_004658.3: p.(Leu3857Val)	missense (23)	22.3-22.4-22.5 (Paternal)	.	.	<i>HERC2</i> (S)
b) Family 105						
NC_000007.14:g. 142771008T>C	NP_001290343.1: p.(Phe9Ser)	missense (19.74)	105.3 (<i>De novo</i>)	rs757333739	0.00008845	<i>PRSS2</i>
NC_000011.10:g.	NP_001287871.1:	missense	105.3	.	.	<i>EMSY</i>

76544332T>C	p.(Met943Thr)	(24.9)	(Paternal)			(2)
NC_000014.9:g.9 3621924A>T	NP_001333147.1: p.(Asn1614Ile)	missense (22.5)	105.3 (Maternal)	.	.	<i>UNC79</i> (2)
NC_000016.10:g. 1205207C>T	NP_066921.2: p.(Pro849Ser)	missense (18.22)	105.3 (Paternal)	rs370675810	0.0000977	<i>CACNA1H</i> (2)
NC_000005.10:g. 66054545G>T	NP_001240626.1: p.(Arg1076Leu)	missense (27.5)	105.3 (Paternal)	rs774967294	0.00000698 3	<i>ERBIN</i> (2)
NC_000001.11:g. 34868751C>T	NP_001073887.1: p.(Arg780His)	missense (22.6)	105.3 (Maternal)	rs745644846	0.00000697 9	<i>DLGAP3</i> (3)
NC_000019.10:g. 1045174G>C	NP_061985.2: p.(Arg463Pro)	missense (22.9)	105.3 (Paternal)	rs3752233	0.0002	<i>ABCA7</i> (3)
NC_000002.12:g. 169259049G>A	NP_004516.2: p.(Ser830Leu)	missense (23)	105.3 (Maternal)	rs767925581	0.000014	<i>LRP2</i> (3)
NC_000002.12:g. 6865123C>T	NP_997198.2: p.(Val192Ile)	missense (16.75)	105.3 (Maternal)	.	.	<i>CMPK2</i> (3)
NC_000021.9:g.3 3797479G>A	NP_003015.2: p.(Glu685Lys)	missense (21.7)	105.3 (Maternal)	rs769546913	.	<i>ITSN1</i> (3)