SAPIENZA
UNIVERSITY OF ROME

Neurodevelopmental genetic drivers of the human 22q11.2 region

PhD school: Biology and Molecular Medicine

PhD course: Human Biology and Medical Genetics - XXXI Cycle

Curriculum: Medical Genetics

PhD Candidate:

Davide Vecchio, MD

Advisors:

Antonio Pizzuti MD, PhD

Viviana Caputo, PhD

A.Y. 2017-2018

TABLE OF CONTENTS

INTRODUCTION

The 22q11.2 genomic region has long been recognized critical for several chromosomal disorders such as the DiGeorge [DGS, OMIM 188400], the Emanuel [der(22)t(11;22), OMIM 609029] and the Cat-Eye Syndromes [CES or invdup(22)(q11), OMIM 188400] whose phenotypes have also been postulated being determined due to either an increased or a decreased amount of several contiguous dosage-sensitive genes. While Emanuel and CES syndromes are rare disorders characterized respectively by 22q11.2 trisomy and tetrasomy, the 22q11.2 rearrangements - which include both the proximal and distal microdeletion and microduplication events [OMIM 188400, 611867 and 608363] - occur more often in general population resulting to date in the most common known genomic disorders. Indeed, if previous studies assessed the DGS prevalence of 1 in 2.000-4.000 live births, recent evidences from several pregnancies' studies have highlighted a cumulative incidence of 22q11.2 microdeletion in 1:992 and, of 22q11.2 microduplication in 1:850 live births (McDonald-McGinn DM et al. 1999, McDonald-McGinn DM et al. 2015, Grati FR et al. 2015, Clements C et al. 2017).

In the last decade, despite substantial progress have been made in understanding the wide spectrum of congenital anomalies and birth defects associated with 22q12.2 segmental aneuploidies, the highest rate of neurodevelopmental disorders' (NDD) comorbidity has been also showed in this cohort compared to any other genetic conditions due to elevated diagnosis of developmental delay/intellectual disability (DD/ID), autism spectrum disorder (ASD), attention deficit/hyperactivity disorder (ADHD) and notably of schizophrenia (Clements C et al. 2017). More in-depth, recent large epidemiologic studies on 22q11.2 deletion syndrome (22q11.2DS) registered ADHD in 37% of children and psychosis in 41% of adults, with an overall elevated ASD risk for both 22q11.2DS and 22q11.2 duplication (22q11.2DupS) as many as 50% of 22q11.2DS and 38% of 22q11.2DupS children received its community diagnosis (Schneider M et al. 2014, Clements C et al. 2017, Hoeffding LK et al. 2017).

These data suggest that the 22q11.2 region could harbour some dosage-sensitive genes whose imbalance may affect the physiologic neuronal circuits' behaviour. Moreover, these genes may not only impinge on 22q11.2 rearrangements' medical history as modifying factors, but they also may be *per se* NDDs' driver-genes.

With this major aim, since deep genotyping of patients carrying different 22q11.2 rearrangements was successfully demonstrated as a valid approach to uncover the genetic basis of several 22q11.2DS co-morbidities and/or previously categorized idiopathic diseases (spanning from congenital malformations to specific psychiatric disorders; Lopez-Rivera E et al. 2017, Greene C et al. 2017) in this work we applied different approaches to further define first, hence to experimentally confirm, the neurodevelopmental pivotal role of this region.

Overall, this research project's results will add new insights on the elucidation of 22q11.2DS/22q11.2DupS neurobehavioral outcomes as well as of NDDs neuroscience physiopathology, widening the scope on those dynamics that constantly shape and re-shape our Central Nervous System (CNS) and whose impairment has been nowadays highlighted as one of the most frequent reasons for medical referral and/or diagnostic workup during the developmental age such to represent a new population health-burden issue. Indeed, NDDs' incidence in Western countries has recently risen up to ~1 in 66 children (Baio J et al. 2018) and, only an early diagnosis has been so far demonstrated able to impinge on lifespan cognitive disabilities assuring those proper access to interventions which avoid further lifetime costs for individuals, families and society (Muhle RA et al. 2018).

CHAPTER 1

22q11.2 neuroscience: bench, bedside and beyond

1.1 Background and state of knowledge in the research field

In the past two decades large-scale genotyping and sequencing efforts have identified several genomic copy number variants (CNVs) greatly expanding our understanding on their contributions to human typical brain and cognitive development (Lin A et al. 2017), as well as their relationship with psychiatric and neurodevelopmental disorders' source (Marshall CR et al 2017; Olsen L et al. 2018). Moreover, probing those naturally occurring reciprocal CNVs in different clinical cohorts highlighted their association with the whole psychiatric spectrum; i.e. those rearrangements occurring at the 1q21.1, 3q29, 15q11.2, 15q13.3, 16p11.2, 16p13.11-p13.2 and/or 17q12 loci have been sequentially described causative for schizophrenia, ASD and/or DD/ID (Sullivan PF et al. 2012, Malhotra D et al 2012, Hiroi N et al. 2013). Others CNVs have been associated with fewer clinical diagnoses, but their clinical characterization is still incomplete to expand the human nosology since additional reports might be added to attempt their correlation, as it was for the 22q11.2 genomic region (Hiroi N et al. 2013).

Indeed, CNVs which map to the chromosome 22q11.2 locus are known by over than 25 years and, they are among the most frequent and clinically best characterised genetic event in humans (McDonald-McGinn DM et al., 2015). It has been also recognized that this repetitiveness in human genome is due to its peculiar genomic structure which shares physical regions of overlap containing copies of chromosome 22-specific low-copy repeats (LCRs) composed of a complex modular structure with high degree of sequence homology (>95%) over large and different in size stretches within the repeats (Portnoï MF et al., 2005). Eight LCR clusters (four centromeric and four telomeric, progressively namely from A- to H-LCR22s, Fig. 1) have been identified at the 22q11.2 region (Descartes M et al. 2008). These LCR22s predispose to homologous recombination events and mediate meiotic nonallelic homologous recombination (NAHR) making the genome prone to generate those rearrangements causative for both 22q11.2DS and 22q11.2DupS (Portnoï et al., 2009; Piccione M et al. 2011, McDonald-McGinn DM et al. 2015). Moreover, it has conventionally accepted that both 22q11.2DS

and 22q11.2DupS can be further categorized in proximal and distal events whereas they encompass those segments respectively flanked between the A- to C- (included) or the D- to H-LCRs22 (Mikhail FM et al., 2014; Burnside RD, 2015).

Figure 1: 22q11.2 genomic region ideogrammatic representation retrieved from UCSC Genome Browser (https://genome.ucsc.edu). 22q11.2 genomic coordinates (chr22:18,640,000-25,080,000), A- to H-LCR22s (highlighted in blue vertical bars) and NCBI RefSeq genes plotted as reported in Mikhail FM et al. 2014 and according to the Human Feb. 2009 (GRCh37/hg19) assembly.

To date, although the pathogenic nature of CNVs on chromosome 22q11.2 and its neurodevelopmental comorbidity have been recognised for several years, the scientific literature lacks certain estimates about incidence, prevalence, disease risks, and diagnostic-outcomes trajectories in general population. For example, if previous studies assessed the DGS prevalence of 1 in 2.000-4.000 live births, recent evidences from several pregnancies' studies highlighted a cumulative incidence of 22q11.2 microdeletion in 1:992 and, of 22q11.2 microduplication in 1:850 live births (McDonald-McGinn DM et al. 1999, McDonald-McGinn DM et al. 2015, Grati FR et al. 2015, Clements C et al. 2017). These data confute those from the only exiting case-control study performed in the Danish population which has recently showed a prevalence of one in 3672 [0.027%; 95% CI 0.012–0.057] for deletions and one in 1606 [0.066%; 0.040–0.107]) for duplications, clearly indicating significative distinct selective pressures on each rearrangement (Olsen L et al. 2018).

The primarily *de novo* 22q11.2DS typically involve 1.5–3 Mb at the 22q11.2 proximal region and, was first identified in patients with DiGeorge syndrome which is characterised by congenital heart defects, recurrent infections, velopharyngeal insufficiency, and facial dysmorphism, often in combination with variable degree of

developmental delay and intellectual disability (Driscoll D. et al. 1993; Olsen L et al. 2018). This event has been recently accounted up to 1% of patients in clinical cohorts with ASD and/or DD/ID and, has been associated with neurological disorders such as epilepsy (Bucan M et al., 2009; Shishido E et al., 2014; Cooper GM, et al. 2011; Sanders SJ et al., 2015; Szatkiewicz JP et al. 2016; Olsen L et al. 2018). By contrast, reciprocal duplications are less frequent than deletions in clinical cohorts of individuals with DD (one in 350) and/or ID (Ou Z et al. 2008; Van Campenhout S et al. 2012; Olsen L et al. 2018). However, according to Olsen et al. 2018, although risk of congenital abnormalities, developmental delay, and intellectual disability is elevated in deletion carriers, the overall prevalence of neuropsychiatric disorders alone is higher in the duplicated ones, which implies that their identification and clinical monitoring should be extended beyond congenital traits and into the developmental-age psychiatry. These evidences have been also variably confirmed from recent large epidemiologic studies on 22q11.2DS cohorts which registered ADHD in 37% of children and psychosis in 41% of adults, with an overall elevated ASD risk for both 22q11.2DS and 22q11.2 as many as 50% of 22q11.2DS and 38% of 22q11.2DupS patients received its community diagnosis (Schneider M et al. 2014, Clements C et al. 2017, Hoeffding LK et al. 2017). Finally, the deletions have been formally associated with high risk and penetrance of schizophrenia (Kirkov et al., 2014). Indeed, one in four individuals born with 22q11.2DS develops schizophrenia, and 0.5–1% of individuals with schizophrenia in the general population carry the associated deletion (Fung WLA et al. 2015). This makes this rearrangement as the most common of clinically relevant schizophrenia-CNVs that combined account for 3.5–5% of its pathogenic causes (Costain G et al. 2013).

Thus, although our understanding of 22q11.2 disease-critical elements is ever-evolving alongside our understanding of the genome (Kragness S et al. 2018), the previous clinical and epidemiologic evidences, taken together, confirm this locus as pivotal for human neurodevelopment as well as a valuable research field for neurobehavioral illness' elucidation whose molecular mechanisms, to date, still remain mostly unknown.

1.2 The 22q11.2 rearrangements signature on human brain

The 22q11.2 region harbours several genes highly conserved in both heterologous and homologous species and, widely brain expressed in ventricular/subventricular

progenitors as well as in mature cortical neurons (Meechan DW et al. 2015). As many of these genes are also early encoded in developing brain, it is still debated how their contiguous impairment can differently compromise their precursors proliferative and neurogenic capacity (Lin A et al. 2017). Indeed, identical rearrangements are often associated with multiple diagnoses and same sized gain- or loss-segmental aneuploidies do not adequately account for the clinical heterogeneity seen in both 22q11.2DS and 22q11.2DupS cohorts that, in some cases, can even get a slight and unpredictable neuro-phenotype (Williams NM et al. 2012; Hiroi N et al. 2013). In past years, different cross-sectional studies showed recurrent neuroanatomical alterations in 22q11.2DS such as: midline defects, increased white matter hyperintensities, and decreased grey matter (Boot E et al. 2012; Van L et al. 2017). More recently, growing numbers of both 22q11.2DS and 22q11.2DupS longitudinal MRI studies have provided the opportunity to further elucidate some anatomical and functional changes that have become promising predictive neuroradiological biomarkers. According to the case-control study performed by Lin A at al. 2017, 22q11.2 gene-dosage impairments conflict with grey/white matter volumes, cortical surface areas expansion and impinge on cortical thickness with opposite developmental pathways. Indeed, a greater cortical thickness reduction is seen in 22q11.2DupS than in 22q11.2DS, while patients carrying deletion events show a smaller brain volume compared to controls and to 22q11.2DupS. These diametric patterns have been showed being extended also into subcortical regions whereas 22q11.2DupS had an average significantly larger right hippocampus, but lower right caudate and corpus callosum volume in comparison to 22q11.2DS (Lin A et al. 2017).

Of note an aberrant frontal connectivity, mostly involving the frontal–frontal, frontal–parietal, and frontal–occipital regions, has been also highlighted in both conditions resembling those impaired networks retrieved in large-scale studies of idiopathic psychotic cohorts (Ottet MC et al. 2013; Mattiaccio LM et al. 2018).

1.3 Known neurobehavioral genes within the critical region

The above described 22q11.2 features constitute a comprehensive neuro-anatomical signature on human brain that, according to the scientific literature, bridges these patients' cohorts with other non-syndromic neuropsychiatric sets. In this view and, emphasizing the idea of the 22q11.2 region as a genomic model for common

complex-trait diseases' elucidation, it is worthy of mention the intense effort supplied by several groups and dedicated international consortia to investigate the 22q11.2 schizophrenia high-relative risk. This approach was in the long run effective, allowing scientists to identify several neurobehavioral high-risk genes and/or their predisposing polymorphisms whose main evidences are individually hereinafter summarized.

The *COMT* gene (OMIM: 116790) is located in proximal 1.5 Mb critical deletion region that is consistently deleted in cases of 22q11.2DS. Its protein product, the catechol-O-methyltransferase, catabolizes several catecholamines such as dopamine, norepinephrine, and epinephrine (Thompson CA et al. 2017). This gene encodes both membrane-bound and soluble COMT enzyme (MB-COMT and S-COMT, respectively), with the former being the predominant form in the brain (Bertocci B et al. 1991; Chen J et al., 2004). The COMT enzyme, along with monoamine oxidase A, has a particularly strong influence on dopamine metabolism, particularly in the prefrontal cortex (Tunbridge EM et al. 2004; Tunbridge EM et al. 2006) where there is a relatively low concentration of dopamine transporters and hence a need for an alternate way to clear dopamine from synapses. Its functional polymorphism at codon 158 of the MB-COMT enzyme, consisting of the substitution of valine with methionine, results in a change from high to low enzymatic activity (Thompson CA et al. 2017). This single nucleotide polymorphism (SNP; rs4680) alters those functions governed by the prefrontal cortex and has been explored in 22q11.2DS with respect to cognition and susceptibility to schizophrenia. However, multiple studies have found no association of the COMT functional Val/Met common allele with schizophrenia in adults with 22q11.2DS (Bassett AS et al. 2007; Goethelf D et al. 2013). However, since results for overall intellect in 22q11.2DS are mixed, there may be some effects of this common variant on frontal lobe functioning and anatomy (Bassett AS et al. 2007; Goethelf D et al. 2013).

The *PRODH* gene (OMIM: 606810) encoding the enzyme proline dehydrogenase, which breaks down proline, has been also extensively studied. In part because pathogenetic mutations in PRODH are known to cause type I hyperprolinaemia, which in severe forms can cause seizures and intellectual disability, in part because alterations in glutamate signalling are very well-established risk factors for psychosis (Moghaddam B & Javitt D, 2012). Indeed, proline is an intermediate in the biosynthesis of glutamate and, proline dehydrogenase catalyses the rate-limiting step in the conversion of proline to glutamate

(Bender HU et al., 2005). Approximately one-third of patients with 22q11.2DS have increased levels of proline, and several studies, although not all, have shown significant associations between high proline levels and various brain outcome measures (Vorstman JA et al., 2009). Noteworthy, studies of common variants in *PRODH*, as for those in *COMT*, show contradictory results with respect to risks for intellectual disability or schizophrenia (Philip N and Bassett A, 2011; McDonald-McGinn DM et al. 2015).

The Zinc-finger DHHC-type-containing 8 (*ZDHHC8*, OMIM: 608784), which encodes a palmitoyltransferase, has shown interesting results in studies of mutant mouse models, with effects on axonal growth and terminal arborization, and potential functional implications for synaptic connections and working memory (Mukai J et al. 2015). Indeed, among 72 known SNPs located within the proximal 1.5-Mb critical region, a polymorphism in *ZDDHC8*, known as rs175174, had the greatest association with schizophrenia (Mukai J et al., 2004). This polymorphism mediates a ZDDHC8 alternative splicing that introduces a premature stop codon into the growing amino acid chain such to terminate translation into a truncated protein with diminished activity (Mukai J et al., 2004; Thompson CA et al., 2017). Another neurodevelopmental candidate is the *RANBP1* gene (OMIM 601180) encoding a binding protein for the small GTPase Ran. As a regulator of the Ran complex, this protein has multiple functions (i.e. cilia formation and modulation of mitosis) that may contribute to the CNS and other phenotypes seen in 22q11.2 rearranged patients. Evidences for a role in neurogenesis also places *RANBP1* as a candidate for both 22q11.2 cortical circuits behavioural impairment and ADHD/ASD pathogenesis (Paronett EM et al., 2014; McDonald-McGinn DM et al. 2015; Clements CC et al. 2017).

The Reticulon 4 receptor, also known as the Nogo-66 receptor (NgR), is a glycosylphosphatidylinositol (GPI)-linked protein encoded by the *RTN4R* gene (OMIM 605566; Maynard TM et al., 2008). Interacting with Nogo-66 (Neurite Outgrowth Inhibitor 66), this receptor plays a significant role in myelin-mediated axonal growth inhibition (Fournier AE et al., 2001). Indeed, Nogo-66 localizes to axons and binds to oligodendrocyte-myelin glycoprotein, myelin-associated glycoprotein, and Nogo-A (neuronal RTN4 isoform), all of which inhibit axonal sprouting. To date some *RTN4R* allelic variations, such as the presence of the G rs701428 allele, were significantly associated with volumetric differences in occipital lobe gray matter and its enhanced

thickness has been robustly associated with those ultrahigh neuroanatomical signs seen in psychotic patients' cohorts (Thompson CA et al., 2017). Moreover while PIK4CA (OMIM 600286), an enzyme that helps to regulate signal transduction in neurons and synaptic transmission, has been related to schizophrenia in both individuals with and without 22q11.2 rearrangements suggesting that it could be *per se* responsible of a non-syndromic psychotic form (Vorstman et al., 2009b), the synaptosomal-associated protein 29 kDa (*SNAP29*, OMIM 604202) has been associated with the cerebral dysgenesis, neuropathy, ichthyosis and palmoplantar keratoderma (CEDNIK), the Kousseff and the Opitz G/BBB syndromes (McDonald-McGinn DM et al. 2013). This is due to its pleiotropic effects; indeed, *SNAP29* encodes a SNARE (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) protein highly expressed in myelinating glial cells, predicted to mediate vesicle fusion at the endoplasmic reticulum or Golgi apparatus membranes and, required for lamellar body formation in the skin, β1-integrin endocytosis and cell migration (McDonald-McGinn DM et al., 2015).

Another gene of interest is *DGCR8* (OMIM 609030), encoding the DGCR8 microprocessor complex subunit (also known as Pasha), a double-stranded RNA-binding protein that mediates the biogenesis of miRNAs (Wang Y et al., 2007). Subtle alterations in miRNA expression levels can have profound effects on brain development and plasticity, especially involving synapses (Beveridge NJ et al., 2010). Recent studies propose that DGCR8 may play a part in modifying the expression of genes outside of the 22q11.2 deletion region that contribute to the neuropsychiatric and other phenotypes associated with (Stark KL et al., 2008; Merico D et al., 2014). Indeed, defective cortical circuitry and some abnormalities of signalling in Sonic Hedgehog and CXC chemokine receptor 4 (Cxcr4)– CXC chemokine ligand 12 (Cxcl12; also known as Sdf1) signalling, which are pivotal in interneuron migration, have been detected in brains of mouse deletion models that may involve a DGCR8-mediated miRNA mechanism and are now hypothesized of relevance for schizophrenia in the general population (Toritsuka M. et al., 2013; Meecan DW et al., 2015).

Finally, dose-dependent expression of CLAUDIN-5 (*CLDN5*, OMIM 602101) was recently described as a 22q11.2-schizophrenia modifying factor. CLAUDIN-5 is expressed in endothelial cells forming part of the blood-brain barrier. Using an inducible knockdown mouse model, Greene C et al. 2017 link Claudin-5 suppression with psychosis through a

distinct behavioural phenotype showing impairments in learning and memory and anxiety-like behaviour. These authors also showed that anti-psychotic medications dose-dependently increase CLAUDIN-5 expression in vitro and in vivo, while an aberrant expression of CLAUDIN-5 in the brains of schizophrenic patients post mortem was observed compared to age-matched controls. Hence, for the first time, this translational research on the 22q11.2 suggests that drugs directly targeting the blood-brain barrier may offer new therapeutic opportunities for treating this disorder in both syndromic and non-syndromic patients' cohorts.

1.4 Deciphering the 22q11.2 neurodevelopmental high-risk

Emphasizing the idea of the 22q11.2 region as a genomic model for complex-traits diseases' elucidation, the lesson learned from the investigation of its schizophrenia-related form has offered a methodological approach to highlight which genes may impinge on a given neurophenotype. Thus, although the understanding of 22q11.2 rearrangements' genetic underpinnings is highly complex (Kragness S et al., 2018), their molecular elucidation rest of pivotal importance to translationally achieve new diagnostic and treatment options such it was recently described for the *CLDN5* gene.

In last decade, despite substantial progress have been made in understanding the wide spectrum of congenital anomalies and birth defects associated with 22q12.2 segmental aneuploidies, the highest NDDs comorbidity rate has been also showed in this patients' cohorts compared to any other known genetic conditions. Indeed, subgroups of ASD, ASD and ID, ID and ADHD, have been progressively identified during the developmental age and, interestingly, mainly reported prior to any psychotic symptoms' onset (Christian SL et al., 2008; Niklasson L et al., 2009; Vorstman JA et al., 2013). These evidences beg the questions of why same-sized genetic variants associate with many clinically distinct disorders and, why some individuals with same rearrangements exhibit not all the disorders but a variable combination of diagnoses. Obviously, there must exist factors that modulate each CNV penetrance and expressivity; however, another possibility is that common mechanisms may be shared by several neuropsychiatric disorders in developmental pathways. This also means that these diseases should not be thought as mechanistically distinct psychiatric entities (Hiroi N et al., 2013).

We agree with this hypothesis which implies both an underlying evolution of neurophenotype impairment through different ages and, a possible common neurobiological substrate. In this view, recent researches are worthy of mention since previous schizophrenia-related genes were also identified as potential ASD-risk mediators in these cohorts. Indeed, Radoeva et al. reported that 22q11.2DS patients with ASD are more likely carriers of both the low-activity alleles of *COMT* and *PRODH* than affected individuals without ASD (Radoeva PD et al., 2014). Moreover, Hidding et al. demonstrated a quantitative relationship between ASD symptoms' severity and combination of COMT-Met genotype plus proline levels in 22q11.2DS (Hidding E et al., 2016). These results suggest that an interaction between COMT and PRODH may increase the ASD risk in these patients and, they confirm an existing bridge among ASD recognizable molecular causes and the schizophrenia-driven ones. This evidence is also supported on clinical side, whose expressiveness is magnified by the fact that many psychiatric symptoms appear or worsen and, above all, merging with age (Kragness S et al., 2018) up to that point it is no longer possible to distinguish confident diagnostic boundaries between single neurodevelopmental comorbidities. Indeed, longitudinal studies have found negative correlations between age, ASD and intelligence quotient (IQ) scores suggesting that these patients may show a gradual decline in their cognitive essay as they grow into adulthood (Gothelf D et al., 2005; Green T et al., 2009). Of note a further severe IQ decline is also observed prior to the onset of schizophrenia (Vorstman JA et al., 2015) and, it has been also documented that neurobehavioural symptoms, such as anxiety, can increase with age (Swillen A and McDonald-McGinn D, 2015).

Thus, based on both above mentioned molecular and clinical features, it is possible to conclude that the neurological impairment trend seen in these cohorts mirrors those observed in idiopathic psychotic diseases. This also suggests that further researches should pinpoint not only the genes that are differentially affected by rearrangements, but also how these genes could be impinged through the advancing age.

CHAPTER 2

A genotype-first approach to the 22q11.2 region

2.1 Scientific hypothesis: research objectives and rationale

Determining the neural substrates of psychotic illness in 22q11DS and 22q11.2DupS has been a major focus of recent research investigations (Sun D et al., 2018). As extensive molecular and clinical studies have accumulated for 22q11.2 CNVs over the past 20 years, in the first chapter we have first summarized those evidences that point out this region neurodevelopmental-driven role. To date, several 22q11.2 genes have been linked to schizophrenia in both individuals with this syndrome and the population at large (Thompson CA et al., 2017) proving an existing causal bridge between these patients' cohorts with other non-syndromic neuropsychiatric sets. Then, we have highlighted epidemiologic measurements, pleiotropic symptoms and traits seen in individuals carrying 22q11.2 CNVs in relation to other neuropsychiatric disorders with early diagnostic onset, such as DD/ID and ASD. However, when we reviewed those attempts made in order to identify any specific 22q11.2 NDDs' causal genes, the literature did not show conclusive results. Consequently, there is still a considerable need in determining which genes contribute to the elevated risk for DD/ID and ASD.

Starting from the evidence that this region harbours some dosage-sensitive genes whose imbalance affects the physiologic neuronal circuits' behaviour, our major research aim was to leverage a novel study design able to uncover new 22q11.2 NDDs' driver genes. Moreover, based on schizophrenia research results, we also assumed that these genes may not only impinge on 22q11.2 rearrangements' medical history as modifying factors, but they also may be *per se* NDDs driver genes in humans.

To achieve this goal, since it has been proposed that a more fruitful strategy to elucidate complex diseases such as DD/ID and ASD should include the examination of intermediate phenotypes (Meyer-Lindenberg A and Weinberger DR, 2006), we applied a "genotype-first approach" to a wide cohort of idiopathic NDDs patients narrowing the investigations to those 22q11.2 brain expressed genes that we highlighted as potential candidates after an extensive enrichment analysis. In this context, intermediate

phenotypes are meant to be biologically-based traits or mechanisms through which genes might affect behaviour, while "a genotype-first approach" constitutes a reverse strategy that assigns the pathogenic effects of many different genes and determines whether particular genotypes manifest as clinically recognizable phenotypes (Stessman HA et al., 2014; Jansen S et al., 2018).

Thus, to address the above-mentioned scientific hypothesis, we outlined and carried out three sequential research steps inclusive of:

I. Enrichment analysis and candidate genes prioritization,
II. NDDs patients' cohort molecular screening by targeted resequencing,
III. Pathogenicity findings and functional background elucidation.

2.2 Enrichment analysis and candidate genes prioritization

To discover which 22q11.2 genes can contribute both to physiologic and, to the previously discussed abnormal neurodevelopment, we first performed a high view analysis of their constitutive brain expression levels assuming that best candidates may lie within this subgroup. The 22q11.2 genomic region harbours ~198 RefSeq annotated genes (O'Leary NA et al., 2015; http://www.ncbi.nlm.nih.gov/RefSeq) and, according to the Genotype Expression database (GTEx, Carithers LJ at al., 2015; https://gtexportal.org), significant transcriptome brain expression data can be retrieved for 83 of them (41.9%, listed in Suppl. Tab. 1). Hence, to further investigate the neurobehavioral role, we stratified their transcripts per million (TPM) values in quartiles assigning for each gene the highest median brain GTEx TPM entry as constant variable (Suppl. Tab. 1). The summary statistics obtained is summarized in Tab. 1.

Table 1: 22q11.2 summary statistics and quartiles distribution cut-off of 121 brain GTEx entries.

Measure	TPM
25% Percentile	0
Median (50% Percentile)	10
75% Percentile	46.5
Mean	28.07
Std. Deviation	44.14
Std. Error of Mean	3.997
Lower 95% CI of mean	20.16
Upper 95% CI of mean	35.99

The so identified 1st quartile of genes consists of 32 elements (overall 16.1%, depicted Fig. 2 and fully listed in Suppl. Tab. 1), whose transcriptome constitutive expression can be variably documented also in several other organs and tissues (Fig. 3).

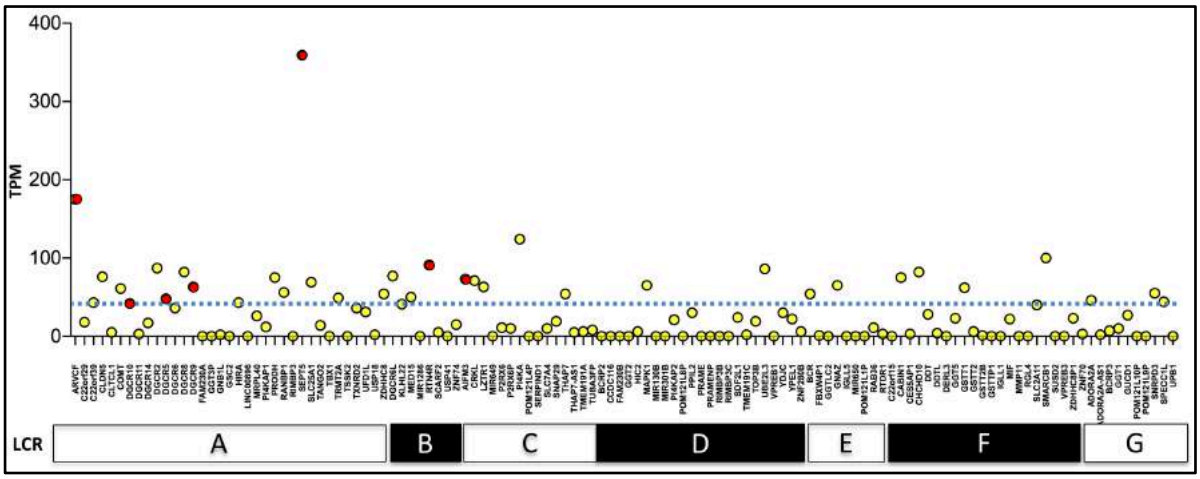


Figure 2: Manhattan plot depicting the genes harboured within the 22q11.2 locus (X axis) and their highest median brain GTEx TPM values (Y axis, yellow dots). Red dots indicate genes that show a brain tissue-specific like expression pattern. 1st quartile brain-expressed genes (also listed in Suppl. Tab. 1) are yellow and red dots plotted above the 75th TPM value percentile (46.5) highlighted by a blue dotted line. Below: white and black boxes part the 22q11.2 region in 7 genomic subsets (namely from "A" to "G") to conventional define segments included between flanking LCRs22. Genomic coordinates, builds and intervals used to perform this ideogrammatic representation are reported in Suppl. Tab. 2.

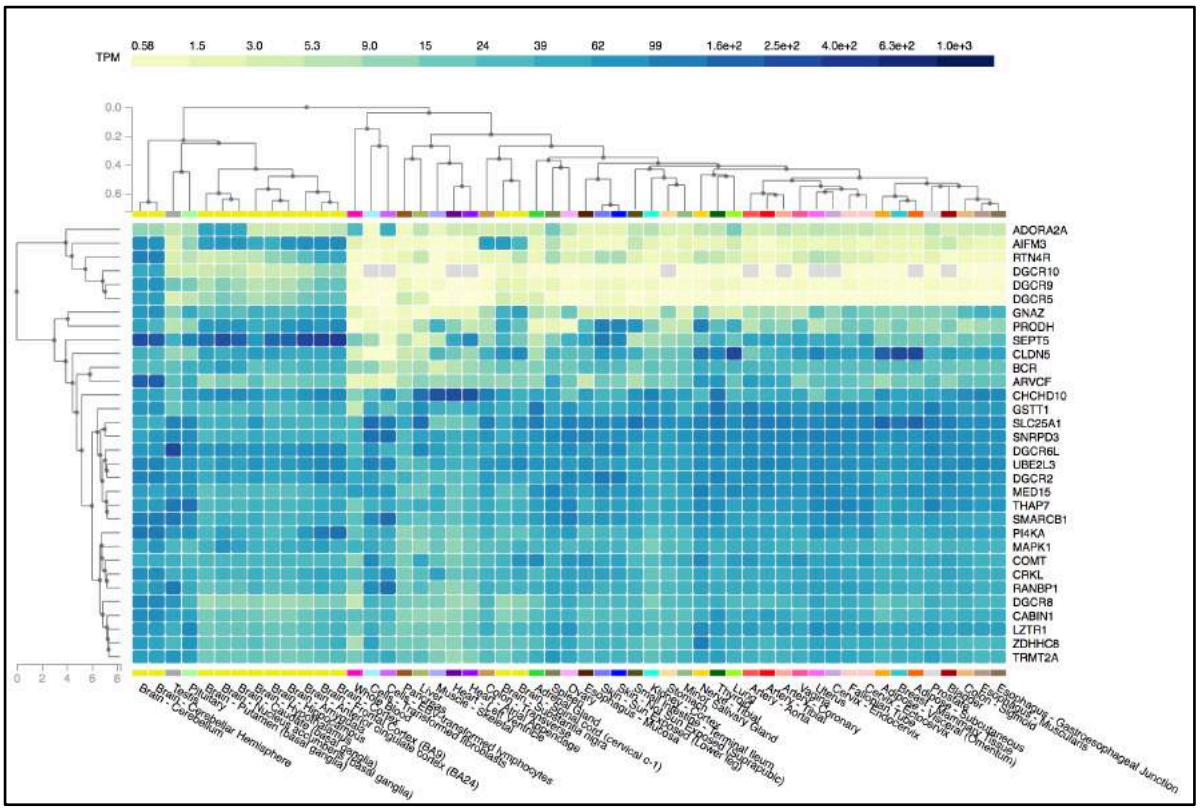


Figure 3: 22q11.2 first quartile brain-expressed genes' set (Y axis, on the right) per all available tissues transcriptome analyses (X axis). Above: TPM-concentration scale intensity plot. Figure generated by GTEx portal (Carithers LJ at al., 2015; https://gtexportal.org) using the "Multi-Gene Query" tool.

Noteworthy, the following 7 genes (overall 3.2%): *SEPT5, ARVCF, AIFM3, RTN4R, DGCR10, DGCR9* and *DGCR5*, showed a brain tissue-specific like expression pattern. This evidence led us to further investigate their neural-type expression. To this end, we used a Cell-type Specific Expression Analysis tool (CSEA; Dougherty JD et al., 2010; Xu X et al. 2014) that, by analysing an RNA profile transcripts' dataset of impaired human brain regions and/or developmental windows, offers an analytical approach to identify a cell population likely to be disrupted according to an inquired genes' list. Interestingly, our set highlighted a statistical significative overlap in habenular nuclei (Fig. 4), which have been recently described acting as CNS regulators connecting forebrain and midbrain with epithalamus and, being involved in behaviour, sleep-wake cycles and learning (Aizawa H et al., 2013; Velasquez KM et al., 2014).

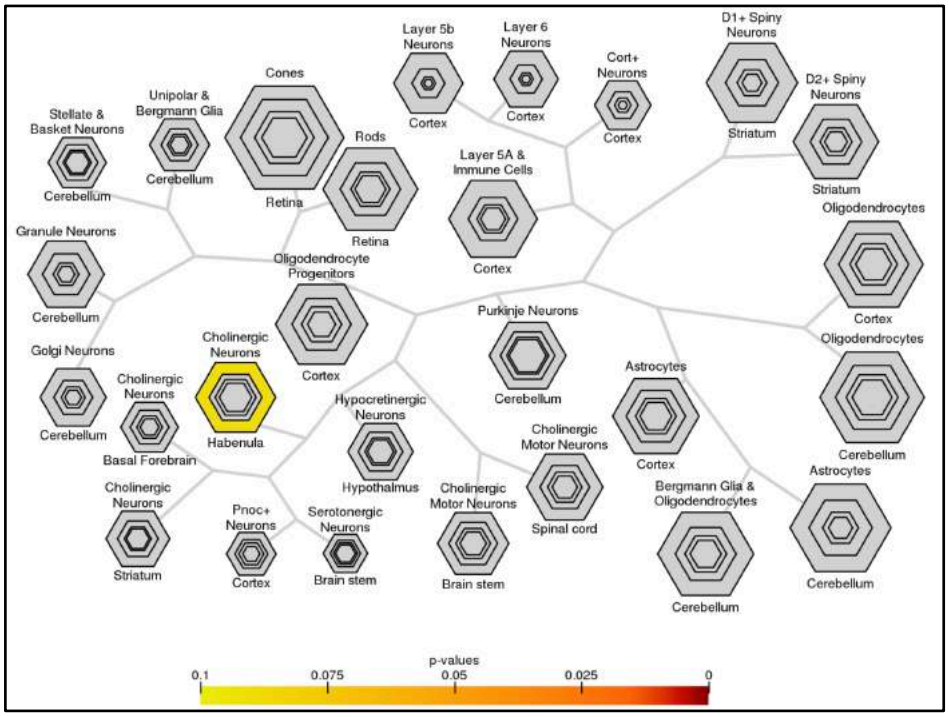


Figure 4: Neuronal-type specific expression analysis' results highlights a statistical significative enrichment overlap in habenular nuclei (p-value 0.043, Fisher's Exact test with Benjamini-Hochberg correction) for 22q11.2 genes which show a brain tissue-specific like expression pattern. In figure hexagons depict each neuronal cell type *per* belonging brain structures hierarchy. The size of each hexagon represents the level of enrichment, where larger hexagons correspond to large enrichment coefficients. Down: p-value scale intensity plot. Figure generated using the CSEA tool (http://genetics.wustl.edu/jdlab/csea-tool-2).

However, since the whole first quartile set of 22q11.2 brain-expressed genes also showed a consistent neural enrichment among all the other cellular components through the Gene Ontology (GO) functional network analysis (Fig. 5), we decided to investigate its hypothesized pathogenic neurodevelopmental contribution both in

general population and, in a defined NDDs patients' cohort. In this view we postulated that this genes' set, compared to all other quartiles taken together, should be outlined carrying a smaller amount of deleterious genetic variants in general population; whereas, it also should be consequently retrieved mostly impinged in NDDs affected patients.

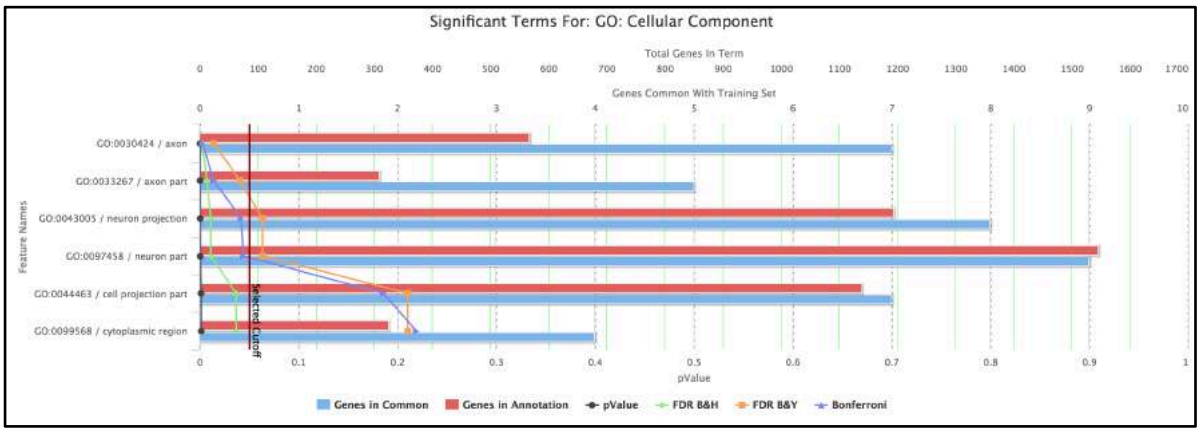


Figure 5: Ideogrammatic representation of the GO cellular component analysis' results showing a neural enrichment (4 out 6 GO terms) for the tested 22q11.2 - 1st quartile brain-expressed genes' set. Figure and p-values (listed in Suppl. Tab. 3) generated by using ToppGenes Suite tool. Abbreviation as follow: FDR, false discovery rate; B&H, Benjamini-Hochberg correction; B&Y, Benjamini-Yekutieli correction.

To address these hypotheses, we collected a 22q11.2 locus-specific genetic deleterious variants' list from both the Exome Aggregation Consortium (ExAC) and the denovo-db v1.6 databases (described in Suppl. Information) exploiting the first as our general population reference dataset (Lek M et al., 2016) and, the second as our probing NDDs patients' cohort (Turner TN et al., 2016). Hence, we summed each deleterious coding variant seen in the 22q11.2 first quartile genes' set versus those retrieved in the other 22q11.2 brain-expressed genes taken together. These two subsets were subsequently analysed and compared in their entirety and, in turn broken out by missense and likely gene-disrupting events (LGD, Turner TN et al., 2016) which include loss of function stop-gained, frame-shift, splice-acceptor and, splice-donor mutations (Suppl. Table 4 and Suppl. Table 5).

The performed statistical analyses confirmed our starting hypothesis (Fig. 6); indeed, we demonstrated that the 22q11.2 first quartile of brain expressed genes is less impinged by deleterious variants in general population (two tailed binomial test p-value <0.001), whereas causing a higher burden in NDDs patients (Wilcoxon rank sum p-value 0.00036) compared to the other 22q11.2 brain-expressed set of genes (overall

counts in Suppl. Table 6). Thus, accounting all the previously stated evidences, we concluded that the 22q11.2 first quartile of brain expressed genes:

1. may exert a downstream functional convergence upon those circuits which mediate both their driver physiological and pathophysiological behaviours;
2. constitutes a reliable set to highlight novel NDDs' candidate genes.

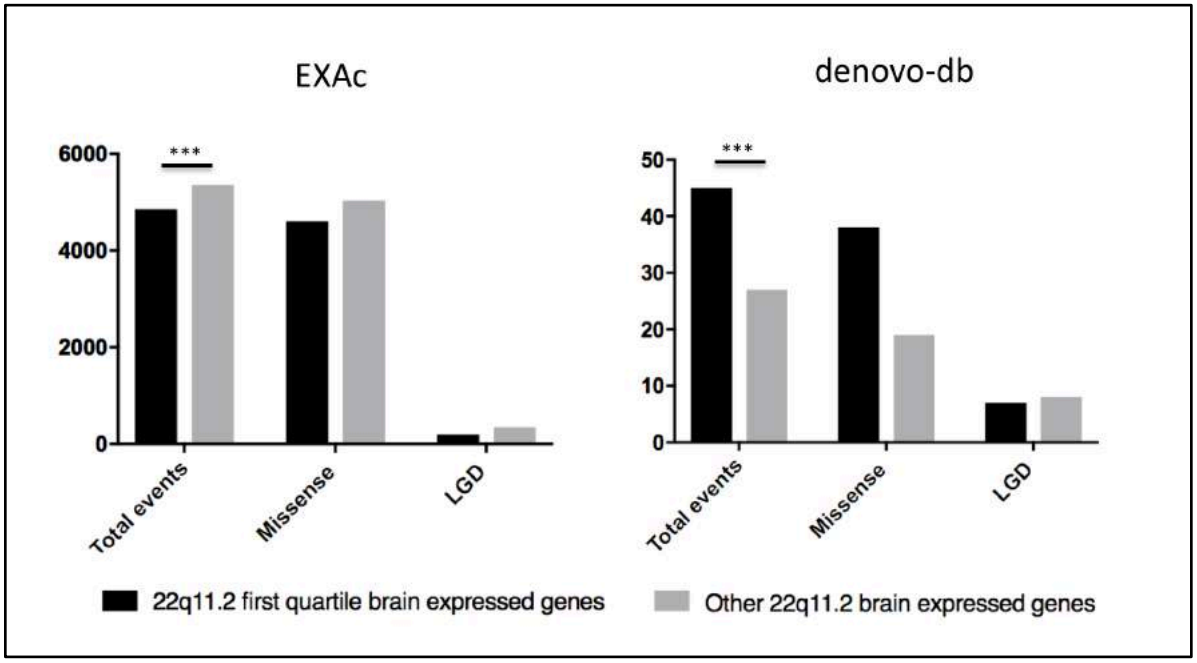


Figure 6: Ideogrammatic representation of the 22q11.2 locus-specific genetic deleterious variants' counts retrieved from EXAc (on the left) and the denovo-db (on the right) databases. Columns depict the overall amount of 22q11.2 first quartile genes' events in black versus those retrieved in the other 22q11.2 brain-expressed genes in grey, both in their entirety (total) and, in turn broken out by missense and LGD columns. Different statistical analysis approaches were performed in these sets using the GraphPad Software, Inc. EXAc values were investigated trough a binomial test comparing the observed variants distribution to expected events. The results (fully reported in Suppl. Table 7) showed a two-tailed p-value of <0.0001 (***, on the left). denovo-db values were investigated by Wilcoxon rank sum test, p-value 0.00036 (***, on the right), because of non-proportional hazard ratios using R package version 1.14.4. Overall, these evidences indicate that 22q11.2 first quartile of brain expressed genes' set is less impinged by deleterious variants in general population, whereas it is also significantly more hit in NDDs patients.

To further increase the statistical power of our enrichment analysis, we used the following selection criteria intersecting this gene set with all of the 22q11.2 genes:

(i) seen carrying at least one missense and/or one LGD event in the denovo-db database (deleterious variants listed in Suppl. Table 8),

(ii) whose loci gain significantly higher rearrangement rates in cases according to a CNV morbidity map generated from the comparison between 29,085 NDD cases and 19,584 controls (Suppl. Notes and Suppl. Table 9; Coe BP et al., 2014),

(iii) whose scored probability rates for any *de novo* deleterious event reach significant values according to at least one of the three models recently developed by Coe BP et al., 2018 in a unique workflow.

The first model incorporates locus-specific transition/transversion/indel rates and chimpanzee-human coding sequence divergence to estimate the number of expected *de novo* mutations (O'Roak BJ et al. 2012), hereafter referred as the chimpanzee-human divergence or CH "model 1". The second model, also known as denovolyzeR and hereafter referred as "model 2", estimates *de novo* mutation rates in the context of 192 trinucleotides. Briefly, in this model the mutability of the central base in a trinucleotide, e.g. ACA → AAA, is measured as the proportion of mutations affecting that base divided by the total count of the trinucleotide spanning the human genome (Ware JS et al., 2015). Finally, while both previous models account for LGD and missense probabilities, the third model ("model 3") described in Coe BP et al., 2018 includes an additional pathogenicity filter that predicts severe missense variants (i.e., CADD scores over 30 or MIS30) prior to applying the CH model 1.

Table 2: Significative statistical results of our 22q11.2 genes' enrichment analysis. In table, models p-values refer to the estimated probability that *de novo* deleterious variants may arise by random chance alone in a given locus broken out *per* LGD or missense (model 1 and 2, formulae in Suppl. Note), and MIS30 events (model 3). All these models apply their underlying mutation rate estimates to generate prior probabilities for observing a specific number and class of *de novo* mutations for a given gene. These models were applied to all the 22q11.2 events retrieved in denovo-db cases and controls probands (Suppl. Table 8). All models were run using R package version 1.14.4 according to the Vv.Aa. reported default settings (O'Roak BJ et al. 2012, Ware JS et al., 2015 and Coe BP et al., 2018). CNV morbidity map p-values are discussed in text and in Suppl. Table 9.

GENE	CNV MORB. MAP_DELS p-value	CNV MORB. MAP_DUPs p-value	Model 1 LGD de novo p-value	Model 2 LGD de novo p-value	Model 1 missense de novo p-value	Model 2 missense de novo p-value	Model 3 Missense CADD30 de novo p-value	Highest CADD score seen in denovo-db	Deleterious variants count in denovo db
ARVCF	8,42E-37	3,42E-09	0,0103	0,0011	1	1	1	42	2
DGCR8	1,56E-36	6,13E-10	0,0472	0,0462	0,3305	0,457	1	10.6	2
TRMT2A	1,28E-40	3,97E-10	1	1	0,5491	0,39	0,02915	32	2
MED15	1,54E-34	3,42E-09	1	1	0,0112	0,0196	1	29	4
CLDN5	1,83E-31	6,03E-12	1	1	0,0287	0,0278	1	29.9	2
RANBP1	7,63E-41	1,18E-08	1	1	2,01E-05	0,127	1	32	1
AIFM3	1,48E-34	4,68E-12	0,124	0,0487	0,2292	0,0777	0,0358	25.6	2
LZTR1	1,48E-34	4,68E-12	0,155	0,0571	0,1226	0,0324	1	40	10

Combining all the previous significative results, we identified a union set of 8 genes: *ARVCF, CLDN5, DGCR8, RANBP1, TRMT2A, MED15, AIFM3* and, *LZTR1* (Tab. 2). Hence, in order to highlight the best candidates, these genes were ranked according to both the number of events and, the CADD score seen in denovo-db database (Fig. 7). Thus, the

LZTR1 and *ARVCF* genes were prioritized respectively due to the associated highest number of deleterious variants and, the highest CADD score retrieved. Indeed, MIS30 are more likely to be functionally equivalent to an LGD mutation and have been shown to be significantly enriched in NDD patients compared to controls (Geisheker MR et al., 2017; Coe BP et al., 2018).

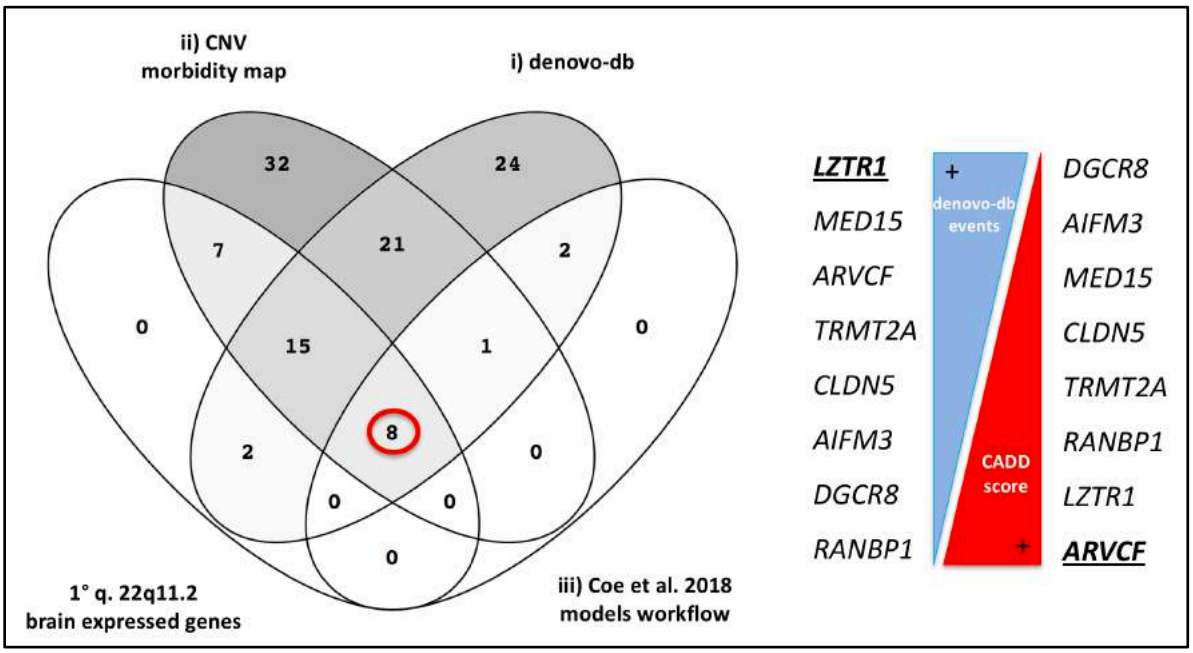


Figure 7: On the left, our enrichment analysis' results depicting the identified intersection set of 8 shared genes (red circle). Whole genes' list *per* intersected group (i, ii, iii) listed in Suppl. Table 10. Candidate genes (on the right) ranked according to both the number of events (blue triangle) and, the CADD score (red triangle) seen in denovo-db database. *LZTR1* and *ARVCF* genes (underlined in bold) were prioritized as best candidates because they harboured the highest number of deleterious *de novo* variants (*LZTR1*) according to denovo-db, or had the highest CADD score of 42 (*ARVCF*) when compared to the other 7 genes.

2.3 NDD patients' cohort molecular screening by targeted resequencing

Large numbers of potentially pathogenic mutations for NDDs have been identified through genetic sequencing, but in most cases only a single occurrence of *de novo* mutation in a particular gene has been discovered (Stessman HA et al., 2017). At the same time, substantially larger numbers of cases and controls are required to explore specific clinical phenotypes. In order to test and verify our enrichment analysis, we targeted resequencing the coding and splicing portions of our two best NDD candidate genes, *ARVCF* and *LZTR1*, in a large cohort of NDD cohorts including ASD, ID, and DD patients by using the molecular inversion probes (MIPs) technology (Stessman HA et al., 2017). The following experiment was carried out at the Eichler laboratory in the

Department of Genome Sciences at University of Washington in Seattle USA as part of a joint research project.

2.3.1 Study materials

MIPs for targeted sequencing

MIPs is a highly sensitive, efficiency and inexpensive approach and perfect for targeted sequencing a group of candidate genes in a large number of samples. MIPs are single stranded DNA molecules around 70 nucleotides, containing a backbone (light grey part) as 30-base linker sequence which common to all MIPs, and targeting arm (dark grey part) as the end sequences which complementary to the target DNA (40 bases in total, most typically around 20 bases for each end hybridization sequence as the universal PCR primer sites). When the probe is hybridized to the genomic target, there is a gap (112 bases) spanning the genomic region of interest to be captured (light blue part). Here, we used an improved MIPs version called single molecular MIPs (smMIPs), which has a molecular tag between the backbone and targeting arm by 8 random nucleotides to enable the assigning of sequencing reads to individual capture events. The schematic of smMIPs is summarized as follow (Fig. 8, modified form Hiatt JB et al.,2013).

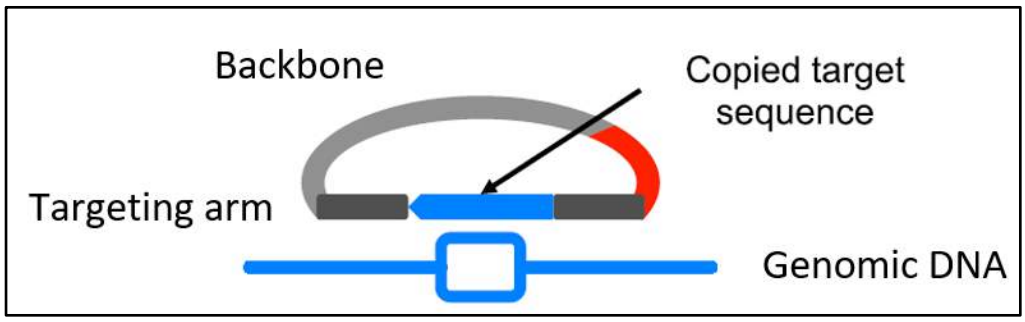


Figure 8: Schematic structure representation of smMIPs.

Samples for targeted sequencing

Whole-blood or cell-line DNA of 11,731 patients with either ASD, ID, or DD diagnoses were collected as part of an international consortium called Autism Spectrum/Intellectual Disability (ASID) network, which involves 15 centers across United States, Belgium, the Netherlands, Sweden, Italy, China, and Australia as previously described in Stessman HA et al., 2017. Only DNA samples from The Autism

Simplex Collection (TASC) and Autism Genetic Resource Exchange (AGRE) cohorts were derived from cell lines. Best-estimate clinical DSM-5 diagnoses were made by experienced, licensed clinicians on the basis of all available information collected during the evaluation. DSM-5 diagnoses included: ASD (299.00), language disorder (315.39), developmental coordination disorder (315.4), attention-deficit/hyperactivity disorders (314.01, 314.00), speech sound disorder (315.39), anxiety disorders (309.21, 300.29, 300.01, 300.02, 300.09), behaviour disorders (313.81, 312.34, 312.81, 312.9), elimination disorders (307.6, 307.7), mood disorders (311.0, 296.99, 300.4), and intellectual disability (319, 315.8). Descriptions, number of individuals represented, and primary ascertainment criteria for each cohort in this study are cumulatively reported in Tab. 3. All experiments carried out on these individuals have been made in accordance with the ethical standards of the responsible institutional and national committees on human experimentation. A proper informed consent was obtained for sequencing, clinical follow-up for inheritance testing, and phenotypic workup.

Table 3: Number of individuals and primary ascertainment criteria for each cohort.

Cohort	Location	Individuals	Ascertainment
ACGC	Changsha, China	1654	ASD
AGRE	NIMH, USA	1732	ASD
San Diego	San Diego, CA, USA	488	ASD
TASC	NIMH, USA	1045	ASD
Adelaide	Adelaide, Australia	2383	ID
Antwerp	Antwerp, Belgium	1089	DD
Leiden	Leiden, The Netherlands	210	DD
SAGE	Seattle, WA, USA	429	DD
Solna	Solna, Sweden	1500	DD
Troina	Troina, Italy	1201	DD
TOTAL		**11731**	

2.3.2 Experimental procedures

smMIPs designing and pooling

smMIPs were designed using MIPgen (http://shendurelab.github.io/MIPGEN/). Each smMIP with unique arms will target a specific genomic region and set to a total fixed length of 162 bp. Five degenerate bases were added between the common linker and the extension arm, which allow a non-duplicate coverage of 4^5=1,024 as the theoretical

maximum. In cases of polymorphisms that may interfere with capture, two smMIPs were designed to capture either haplotype. smMIPs were designed against the GRCh37 human genome reference using dbSNP138. Finally, we designed 43 smMIPs for ARVCF and 37 smMIPs for LZTR1 to cover the coding and splicing (5 bps) regions for the whole gene. Oligonucleotides (Table 4) were ordered from Integrated DNA Technologies (IDT, Coralville, IA). smMIPs were pooled together by gene. For initial testing (1X pool), probes were combined at equal molar concentrations and phosphorylated. After initial testing, probes that performed poorly were repooled and phosphorylated with increased molar ratios at 10X or 50X to rescue the capture coverage. The final pools (1X, 10X, and 50X) were combined as a working pool. We tested and rebalanced each pool independently using 16 unaffected (HapMap) samples as controls.

Multiplex capture and PCR amplification.

For capture experiment, we used 100 ng of genomic DNA to hybridize with 1X Ampligase buffer (Epicentre, Madison, WI), 0.32 μM dNTPs, 0.5 × of HemoKlenTaq (0.32 μl; New England Biolabs, Inc., Ipswich, MA), one unit of Ampligase (Epicentre) and MIPs in a 25 μl reaction. Gap filling and ligation were also performed in this reaction. The amount of smMIPs needed was based on the 1X pool concentration on a ratio of 800 smMIPs copies to each haploid genome copy. Reactions were incubated at 95°C for 10 min and then 60°C for 22h for capture step. Then 2 μl of exonuclease mix were used to degrade linear DNA for incubation at 37°C for 45 min then 95°C for 2 min for exo treatment. For PCR amplification of the captured DNA, 5 μl of ~96 different barcoded libraries were pooled and purified with 0.9X AMPure XP beads (Beckman Coulter, Brea, CA) according to standard protocol. 100 μl of 1X EB (Qiagen, Valencia, CA) were used to resuspend the libraries. Then gel visualization with 2% nondenaturing polyacrylamide gel and quantification with the Qubit dsDNA HS Assay (Life Technologies, Grand Island, NY) was performed according to the manufacturer's protocol. In terms of enough read depth, we pooled ~192 individuals as the a megapool from multiple libraries for sequencing on one lane of Illumina HiSeq 2000 and 101bp paired-end reads were generated according to the manufacturer's protocol.

Table 4: smMIPs designed for ARVCF and LZTR1.

MIP name	Chr	Target start	Target end	MIP sequence	Relative Concentration	Success Rate*
ARVCF_01	22	19958724	19958885	GCGCACGGGTGGGCATTAGAGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGATGAGAGAACGTACCAGG	1x	0.8542
ARVCF_02	22	19959394	19959555	GCTCTGAGGGAGGCCCTGTAGGAGCAGATCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCACATGACCACTTACT	1x	0.7292
ARVCF_03	22	19959275	19959436	GCTGGGTTGTGGGGCCAAGCCCAGGAACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGAGGCCTCTGAGAAG	1x	0.5729
ARVCF_04	22	19959843	19960004	GCAGTGTGGTTGGTTTTCTGTGCCGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGTGTGGGCTTCTCTG	1x	0.8281
ARVCF_05	22	19960212	19960373	CCTAGAGGGTGGCGACCCCAGCTGCCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGTGTCTGTGTTTTATG	1x	0.6719
ARVCF_06	22	19960426	19960587	GCCGGGGGCGGGGTCAGTGGTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGGAAGGCCAGGGAAGCCG	10x	0.6979
ARVCF_07	22	19960611	19960772	GCGGGGTCGGTGGCTTGGGGACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGCCTGGAGGAAGACACCG	10x	0.0052
ARVCF_08	22	19960723	19960884	GTGGCGGTGCTCAACACCATCCACGAAATCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGGTGGGGCAGAGGTG	10x	1.0000
ARVCF_09	22	19961213	19961374	CGACAAGGTGGTGCGCGCCGTCGCCATCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGGAGGTGGGAGAGTG	1x	0.7135
ARVCF_10	22	19961108	19961269	GCACCGAACTGTGCCTGCAGCTTTGGGGTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGTGCTTGTGGAACTG	1x	0.7656
ARVCF_11	22	19961687	19961848	TTCCCGCTTCAGAGCTCCCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGTGTTGAAGTTCCGGCTCTCCGTG	1x	0.3490
ARVCF_12	22	19961567	19961728	GAGGGAATCCCAGGCACCTCACTGGTGCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTCTCTACCTCTCCCTC	1x	0.5990
ARVCF_13	22	19963136	19963297	GCTCAGGATTGCCCTGGTAGTGAAGCTTTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCCACCTTTCCCTTCCA	1x	0.8333
ARVCF_14	22	19964119	19964280	GCCCTGCCCCACCCCTGGCTTCATCACTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTCTCCTCTCTCCTCTGC	1x	0.7448
ARVCF_15	22	19965019	19965180	GCCGACAGGTACCAGGAGGCCGAGCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGGCCAGTCTTTCCTG	1x	0.7292
ARVCF_16	22	19964944	19965105	GTTCCGCATGATGCACACGCAGTTCTCCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCCTTCTTGCCTCCAAA	1x	0.7188
ARVCF_17	22	19964851	19965012	GCTCTCTGTACACCCACAGCAGCCCATCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCCCTGGGCAGTGCTGT	10x	1.0000
ARVCF_18	22	19965482	19965643	GCTGTGGGCCGGAAGGACACTGACAACAACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGTTTCTGGTAGCCAGG	1x	0.8281
ARVCF_19	22	19965437	19965598	GCCTCAGCACCATCGGAGCTCACATTCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGTCTCTCCAGCCCTTG	1x	0.7448
ARVCF_20	22	19966474	19966635	GCGATTGGCCAATCTGTGCTGACCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTTGGAGTCCTCGTTGGGCTC	1x	0.3281
ARVCF_21	22	19966379	19966540	GCACGATCACCTCGTGGGTCAGCGTCTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCTGGTGTGCATGTGGG	1x	0.8698
ARVCF_22	22	19967597	19967758	GATAGCGCCCGCAAGGAGCCGCGCTGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGAGGACACAGCAGATG	10x	0.7292
ARVCF_23	22	19967482	19967643	TCCACTGAGGGCGAGCGCCGCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGCTTGACACCCTCGTTCTCAA	10x	0.2292
ARVCF_24	22	19967485	19967646	GCTTTGAGAACGAGGGTGTCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGGTGCGGCGCTCGCCCTCAGT	1x	0.0469
ARVCF_25	22	19967257	19967418	GCTTGTCACTGGTGAGTGGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGGTGCGGCGCCGGGCCTGT	1x	0.2135
ARVCF_26	22	19967134	19967295	CTCACGGACCTCGTTGTCCCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTCTCAGCAGCTGCACTCCTGCCTAC	1x	0.3333
ARVCF_27	22	19967427	19967588	GCTTGTGGCACTGCTGGACCACCCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGAGCTGCCTGAGGTG	1x	0.5000
ARVCF_28	22	19967134	19967295	GTAGGCAGGAGTGCAGCTGCTGAGACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGGACAACGAGGTCCGTGA	1x	0.5000

ARVCF_29	22	19967743	19967904	GCTGAGAGCACCTGTATCTTGGCCCGGAGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCATCTGCTGTGTCCTC	1x	0.5000
ARVCF_30	22	19967708	19967869	GCGAGCTGGCGGACGAGCGGCCTGCGTTCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCTGCTGGGTGAGTTTG	1x	0.0677
ARVCF_31	22	19967361	19967522	GCAACCTCTCCTATGGCCGCGACACTGACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCTACCTGCAGCATCTG	1x	0.5260
ARVCF_32	22	19968828	19968989	GCTTGGAGGATGACACGCGCAGCCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGTGCTGGCCCCCTTGG	1x	0.0573
ARVCF_33	22	19969044	19969205	GCAGGGGGAGTCCATCTGGGCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGGGCTCGGGGCCTTCGGG	10x	0.4427
ARVCF_34	22	19968941	19969102	GCTGAGGTAGGCTCGAGAGAGTGTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGCAGTGTGAAGCAGC	1x	0.2760
ARVCF_35	22	19968756	19968917	GCACGGCCACAAGGAGGAGGCCTGAGTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGGGTCCTGAGCCTGG	1x	0.1615
ARVCF_36	22	19968671	19968832	GCCACCCTCGTCATCAGCGGCCAGGCTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGTTGGGGGGTGGAGG	1x	0.6771
ARVCF_37	22	19969127	19969288	GTGGATGGATAGGCAGGTAGGTGGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGGAAATGCCGGTCCAGG	1x	0.6563
ARVCF_38	22	19969422	19969583	GCACATCAGGTGCCTCCGGCATCGTGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGCTCACCCACCCACT	1x	0.1042
ARVCF_39	22	19969473	19969634	GCTCCTGGGGCAAGGAGGGACATTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGGTTGTGCCATCTTCGGATG	1x	0.7292
ARVCF_40	22	19969532	19969693	GCTGGAGGAGACCGTGACGGTGGAGGACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGAGCTGCTGTGTGTGG	1x	0.6615
ARVCF_41	22	19978161	19978322	GCCCAGCAGCCTGGCATGGTCAGTGGTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGGTCATGGAGGACTG	1x	0.7292
ARVCF_42	22	19978049	19978210	GCCAAGTGGGAAAACAATCAGTGGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGTTGCCCTACAGCTGGAG	1x	0.8698
ARVCF_43	22	19978256	19978417	CGGTGAAGGAGCAGGAGGCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGTGAGTGGGGCGGGCAGGGG	1x	0.6354
LZTR1_01	22	21336571	21336732	GCAGGCGGCGCGCGGTCCAAGGTACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGCTCGCCGGGAAATGTGG	1x	0.8490
LZTR1_02	22	21336688	21336849	AGGGCCGCAGCCCCGATCTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGAACTCGTCGCAGGGCGGGAG	10x	0.4688
LZTR1_03	22	21336736	21336897	GCAGCTATGGTCGAAGTCCACGCTCGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCCTTCCTGTCCTCAGG	1x	0.2396
LZTR1_04	22	21337297	21337458	GTCAGGGACTGGGCCTGTCCAGCTCCATTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTCCACTCCTTTCTTTC	1x	0.8906
LZTR1_05	22	21340101	21340262	GCACAGTGCTGGGTCCCAGTTGCTAGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGCCACCCTCTCTTCC	1x	0.9063
LZTR1_06	22	21341739	21341900	GCGGGGGTCACCCATGAGGTCCACCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGCCCCTGGGAGGCAAGAG	10x	0.7656
LZTR1_07	22	21342275	21342436	GTGAGAAACTTTGCAGAAACATTTGGGACCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCAGTTTCTCACTCTCT	1x	0.7760
LZTR1_08	22	21343050	21343211	ACCCCACCTCCGACAGCACTGAGACCCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCTCCCTCCCCTCTTCCCT	1x	0.7031
LZTR1_09	22	21343869	21344030	GTGAGGGGCACGGGGAGCCAGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGACCACCAGACCCAAGGG	1x	0.0104
LZTR1_10	22	21344620	21344781	CCGGCCTCACAGGGCCCGCTCACATTTCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTTGGTTATTTTGGCTC	1x	0.7969
LZTR1_11	22	21344742	21344903	GCCTCTTGCACCTGGTGGCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGTTTGTATTCTCTGGGCAAAGCG	1x	0.8958
LZTR1_12	22	21345926	21346087	GCCCCGGAGCAGGTGTTCAGTTGGGATGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCAGGTCTGGAAGTCCA	1x	0.4271
LZTR1_13	22	21345868	21346029	ATCAATGCTGGGACAGGGGCTCCTGCTCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCACCCCCAAACACATA	10x	0.8333
LZTR1_14	22	21346014	21346175	GCCAGCTGGACACCAGTAGCTCCTACCCTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTATGTGTTTGGGGGTG	1x	0.0052
LZTR1_15	22	21346454	21346615	GCATGAGACCAGTTGGCAAGGGCCAGGTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAAGTCCAGGCCAAACA	1x	0.8125
LZTR1_16	22	21346535	21346696	GTCAGGGTGGGCACCTCCTCGGAAGCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGGGGTGATGGAGTCT	1x	0.5521
LZTR1_17	22	21347067	21347228	GCCTGTGGGCCTGTAGAGCCGGCTGGGTCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCCTTCTTGTCCCCCAG	1x	0.2396
LZTR1_18	22	21347990	21348151	GTGGTGCTGACCTTGGCTGGCTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGGCGGCTGTGGGAGAG	1x	0.8646

LZTR1_19	22	21347889	21348050	GTTCGTGCTGGGTGAGGTGGGTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNACAGCCACACTGGGGCCA	1x	0.8333
LZTR1_20	22	21348136	21348297	ACCCAGCCAGCCAAGGTCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTCTCCCGCGCCTGCGTGATCTTCCT	1x	0.7865
LZTR1_21	22	21348233	21348394	GCTCCGCGCTGTGACAATGGCTACGTGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTTCTGTGGGGAGAAGG	1x	0.7240
LZTR1_22	22	21348376	21348537	GCTCATGCAGTTCCTCTACACCGACACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNCTCCCTTCTCCCCACAG	1x	0.6302
LZTR1_23	22	21348497	21348658	GAGCACCTCGAAGGGCCGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGGGCCCATCAGTAAGGCAGG	1x	0.6302
LZTR1_24	22	21348480	21348641	GCCAAAAGGTGGGTGCTGCCAGCCCTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNATCCGGGAGGCCGAGG	1x	0.0156
LZTR1_25	22	21348874	21349035	GTGTGGGGTGGGGTCAGCGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGGATGTGTACAAACTGGCAC	1x	0.8802
LZTR1_26	22	21348826	21348987	GCAGAACGTGCTGGTTGTGTGCGAGAGTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTTCGGGGGCTCTGGGG	1x	0.5313
LZTR1_27	22	21349201	21349362	CGCTCGAACTCCTTCATCATGATCACCTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAACACCATCCCTCCCA	1x	0.3490
LZTR1_28	22	21349134	21349295	GGGGCAGACAGGCACAGGCAGGCCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAAGGGAGTGCGAGGGGG	1x	0.8281
LZTR1_29	22	21350055	21350216	GGGCCTGGATGGTGTCTTCGTTCTGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNATGAAGGCATACCTGGAGGG	10x	0.8281
LZTR1_30	22	21349983	21350144	GCCTGGGATTCTGGCCTTGGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGATAGCCTTGTGGGCTGGCCGTG	1x	0.6510
LZTR1_31	22	21350229	21350390	GCTGCAGACAACGAGCGGTAGGCACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGGGCGGCATGTTGACCT	1x	0.8490
LZTR1_32	22	21350275	21350436	GTTCACCTGCCCATCTTCGGGCATGAAGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGGGGACCTGGGAGTT	10x	0.8490
LZTR1_33	22	21350205	21350366	CCTGGCCGTCAGCAGAACGAAGACCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNGTAGATGTAGCGCAGCATGG	1x	0.6823
LZTR1_34	22	21350952	21351113	GCCTGGTTAAGGAGCTGACTAAAGGAAGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTGTGCACGGGGCTGGGG	1x	0.5729
LZTR1_35	22	21351123	21351284	GCTGCCCCTTTCACCCACTCAGTGGGAGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNAGTCCTGAAGGGGAGG	1x	0.7240
LZTR1_36	22	21351505	21351666	GCAACTTGGAGACCTGTAAGGCAAGCAGGCTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTCTTCACAATGGGCAG	1x	0.8854
LZTR1_37	22	21351476	21351637	GGAGATGGGACCAGATCCACCTACTTCAGCTTCCCGATATCCGACGGTAGTGTNNNNNTCAGATGTCGGCGCCCAGCTCT	1x	0.7240

smMIP sequencing data analysis.

Sequencing reads were analyzed as previously described (Hiatt JB et al., 2013). For primary single-nucleotide variation, indels, and target coverage analysis, initial 101 bp paired-end reads were trimmed to 81 bp before mapping. Sequences were mapped to hg19 using BWA-MEM. Sample-tag indices were made using the 5bp degenerated sequence, which was removed from the beginning of read2 and added to the index barcode. For QC, read-pairs with incorrect pairs and insert sizes were removed after mapping, leaving only the reads with the highest quality scores. We applied FreeBayes for variant calling, and variants were filtered based on read depth > 8 (DP >8) and read quality > 20 (QUAL > 20) in each megapool. Then all variants were filtered against common variants using dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP). We applied a frequency filter of "allele count (AC) <3", and annotated variants on the longest isoform of each gene using SeattleSeq138 and the NCBI 37/hg19 reference.

Sanger validation for variant.

Variants were validated with PCR and Sanger sequencing. Primers were designed using Batch Primer 3 by uploading sequence files in FASTA format according to the user manual with the optimal PCR product size set to 300 bp. For picked primers, we performed BLAT and in silico PCR on the UCSC Genome Browser to avoid multiple hits. PCR was performed with a standard program in 25 µl reaction volume. PCR products were purified with 0.9X AMPure XP beads and the product was visualized on a gel before sending to Sanger sequencing. These variants were annotated with the Ensemble Variant Effect Predictor tool for GRCh37 and with CADD scores (Kircher M et al. 2014). All private LGD variants and MIS30 (missense variants with a CADD score >30) variants were validated by Sanger sequencing. Specifically, a MIS30 was chosen for validation, because these events are very rare (<0.1% of all missense events in control genomes) and are more likely to be pathogenic (Wang T et al., 2016).

2.3.3 Results

After quality control, we validated 38 (LGD/MIS30) QC-passing variants excluding those reported in ExAC and in dbSNP (Tab. 5). Of these, we identified 26 novel LGD and

MIS30 variants for the *LZTR1* (Fig. 9a) and 12 for *ARVCF* (Fig. 9b). To verify the DN status, PCR was also performed on parental DNA where available.

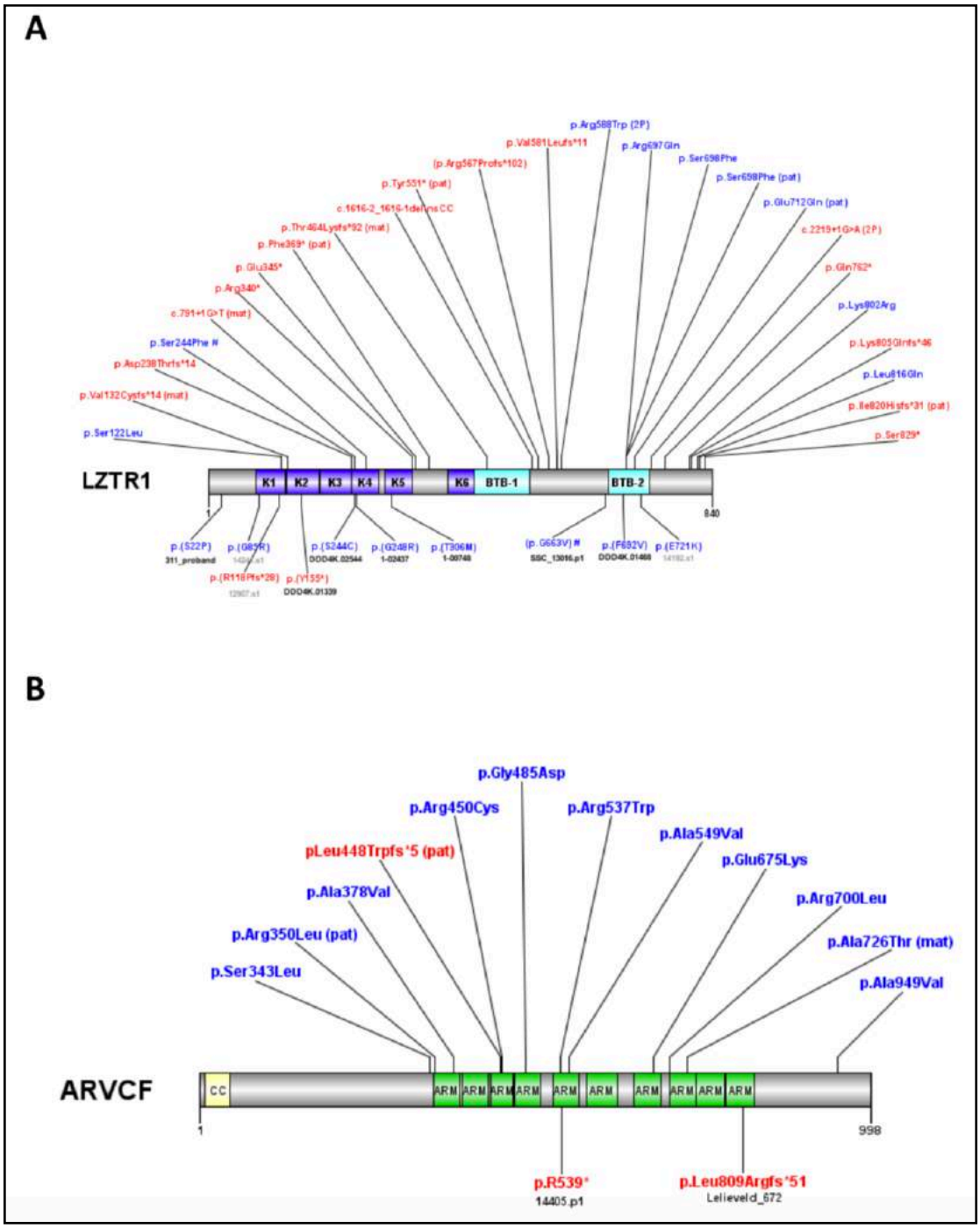
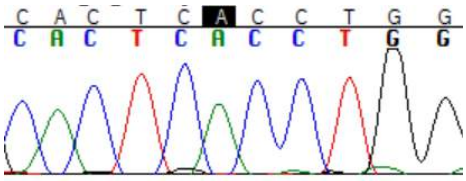


Figure 9: Protein locations of disruptive variants in new candidate NDD risk genes. (A-B) Protein diagrams of LZRT1 (A), and ARVCF (B), with novel LGD (red) and MIS30 (blu) mutations identified in this study and, indicated according to Human Genome Variation Society format. Annotated protein domains are shown (colored blocks) for the largest protein isoforms. Previously reported variants in denovo-db (annotated below the protein structure) are compared with new variants in this study (top). Domain abbreviations: K, kelch; BTB, broad-complex - tramtrack and bric a brac; CC, coiled coil; ARM, armadillo repeat. In brackets: mat, maternal inheritance; pat, paternal inheritance; 2P, two probands.

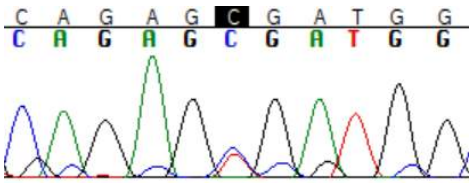
Table 5: Novel LGD and MIS30 mutations identified in this study and their available Sanger traces.

Variant ID	CHR	POS	REF	ALT	Sanger	CADD	GVS	HGVS
ARVCF_var_01	22	19958794	G	A	het	31	missense	p.Ala949Val
ARVCF_var_02	22	19961211	C	T	maternal	32	missense	p.Ala726Thr
ARVCF_var_03	22	19961288	C	A	het	32	missense	p.Arg700Leu
ARVCF_var_04	22	19961682	C	T	het	34	missense	p.Glu675Lys
ARVCF_var_05	22	19965533	G	A	het	32	missense	p.Ala549Val
ARVCF_var_06	22	19965570	G	A	het	34	missense	p.Arg537Trp
ARVCF_var_07	22	19966546	C	T	het	32	missense	p.Gly485Asp
ARVCF_var_08	22	19967314	G	A	het	34	missense	p.Arg450Cys
ARVCF_var_09	22	19967319	AG	A	paternal	31	frameshift	p.Leu448Trpfs*5
ARVCF_var_10	22	19967529	G	A	het	33	missense	p.Ala378Val
ARVCF_var_11	22	19967613	C	A	paternal	32	missense	p.Arg350Leu
ARVCF_var_12	22	19967634	G	A	het	33	missense	p.Ser343Leu
LZTR1_var_01	22	21341837	C	T	het	35	missense	p.Ser122Leu
LZTR1_var_02	22	21341862	G	GT	maternal	35	frameshift	p.Val132Cysfs*14
LZTR1_var_03	22	21344732	CG	C	het	34	frameshift	p.Asp238Thrfs*14
LZTR1_var_04	22	21344754	C	T	de novo	33	missense	p.Ser244Phe
LZTR1_var_05	22	21344815	G	T	maternal	20.7	splice-donor	c.791+1G>T
LZTR1_var_06	22	21346527	C	T	het	38	stop-gained	p.Arg340*
LZTR1_var_07	22	21346542	G	T	het	37	stop-gained	p.Glu345*
LZTR1_var_08	22	21346614	TTTGGCACCACCTC	T	paternal	34	frameshift	p.Phe369*
LZTR1_var_09	22	21348249	AC	A	maternal	35	frameshift	p.Thr464Lysfs*92
LZTR1_var_10	22	21348845	AG	CC	het	24.8	splice-acceptor	c.1616-2_1616-1 delinsCC
LZTR1_var_11	22	21348928	G	GC	het	35	frameshift	p.Arg567Profs*102
LZTR1_var_12	22	21348970	TG	T	het	32	frameshift	p.Val581Leufs*11
LZTR1_var_13	22	21348993	C	T	het	31	missense	p.Arg588Trp
LZTR1_var_14	22	21348993	C	T	het	31	missense	p.Arg588Trp
LZTR1_var_15	22	21350272	G	A	het	37	missense	p.Arg697Gln
LZTR1_var_16	22	21350275	C	T	paternal	33	missense	p.Ser698Phe
LZTR1_var_17	22	21350275	C	T	het	33	missense	p.Ser698Phe
LZTR1_var_18	22	21350316	G	C	paternal	34	missense	p.Glu712Gln
LZTR1_var_19	22	21350402	G	A	het	15.4	splice-donor	c.2219+1G>A
LZTR1_var_20	22	21350402	G	A	het	15.4	splice-donor	c.2219+1G>A
LZTR1_var_21	22	21351049	C	T	het	45	stop-gained	p.Gln762*
LZTR1_var_22	22	21351254	A	G	het	35	missense-near-splice	p.Lys802Arg
LZTR1_var_23	22	21351524	T	TC	het	35	frameshift	p.Lys805Glnfs*46
LZTR1_var_24	22	21351561	T	A	het	30	missense	p.Leu816Gln
LZTR1_var_25	22	21351570	A	AC	paternal	35	frameshift	p.Ile820Hisfs*31
LZTR1_var_26	22	21351600	C	G	het	45	stop-gained	p.Ser829*

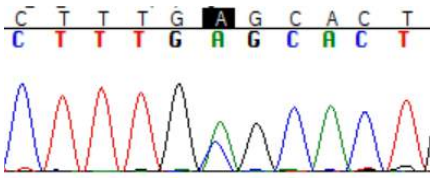
Available Sanger traces for ARVCF variants



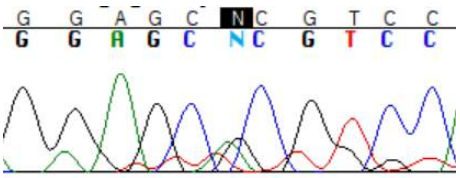
ARVCF_var_01: G>A



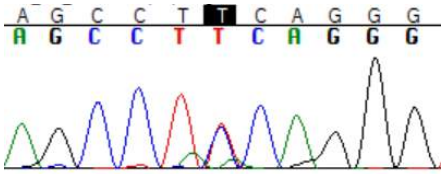
ARVCF_var_02:C>T



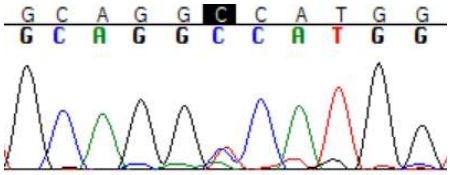
ARVCF_var_03: C>A



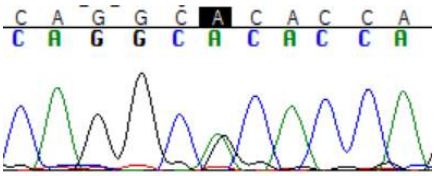
ARVCF_var_04: C>T



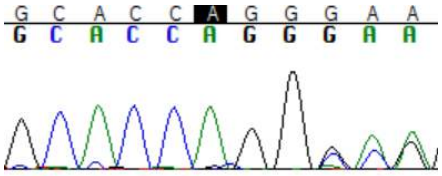
ARVCF_var_05: C>T



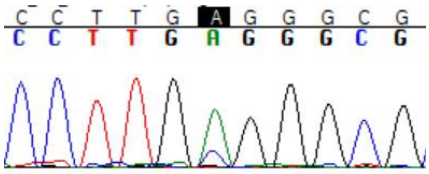
ARVCF_var_07: C>T



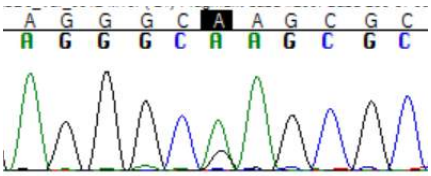
ARVCF_var_08: G>A



ARVCF_var_09: AG>A

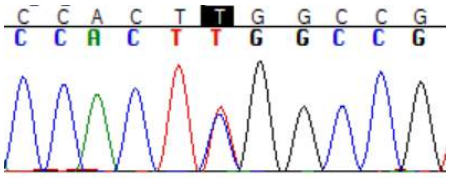


ARVCF_var_11: C>A

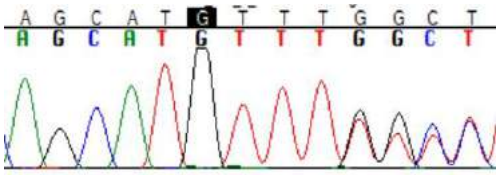


ARVCF_var_12: G>A

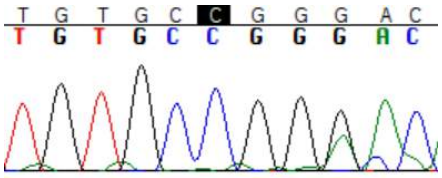
Available Sanger traces for LZTR1 variants



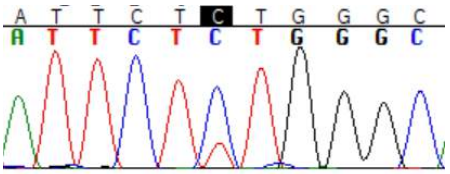
LZTR1_var_01:C>T



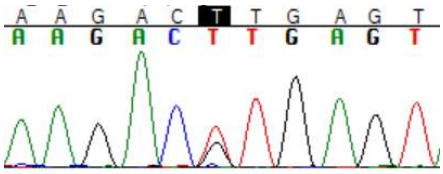
LZTR1_var_2: G>GT



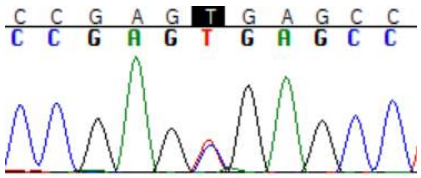
LZTR1_var_03:CG>C



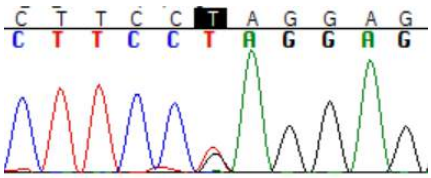
LZTR1_var_04:C>T



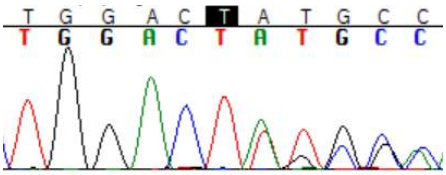
LZTR1_var_05:G>T



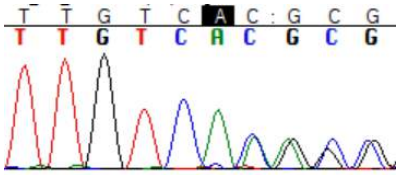
LZTR1_var_06:C>T



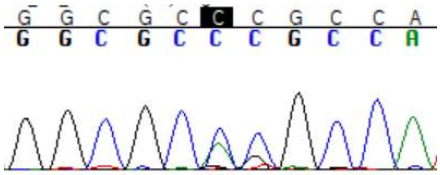
LZTR1_var_07:G>T



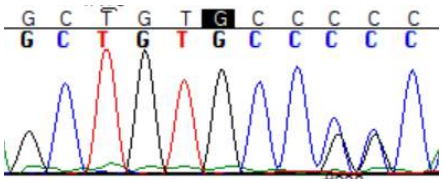
LZTR1_var_08: TTTGGCACCACCTC>T



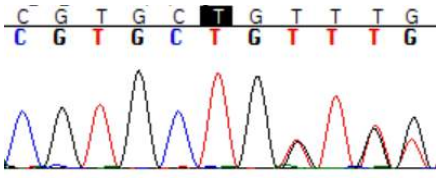
LZTR1_var_09:AC>C



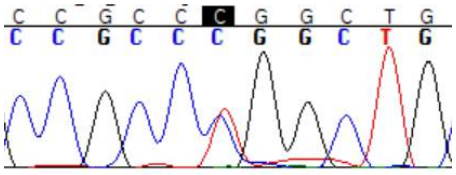
LZTR1_var_10:AG>CC



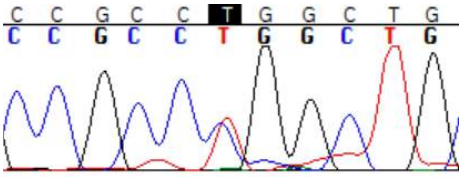
LZTR1_var_11:G>GC



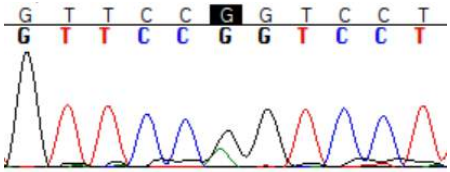
LZTR1_var_12:TG>T



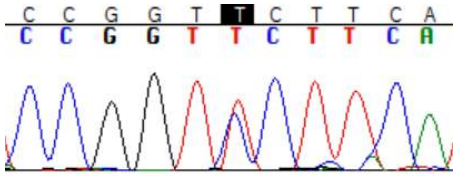
LZTR1_var_13:C>T



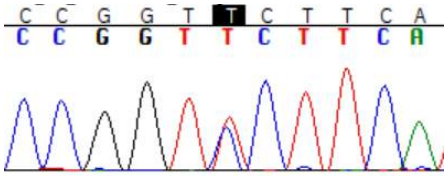
LZTR1_var_14:C>T



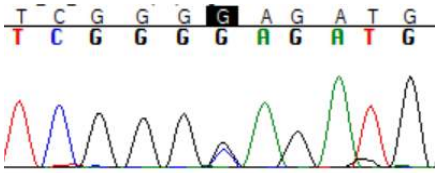
LZTR1_var_15:G>A



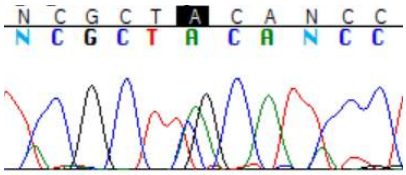
LZTR1_var_16:C>T



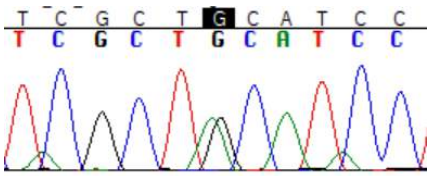
LZTR1_var_17:C>T



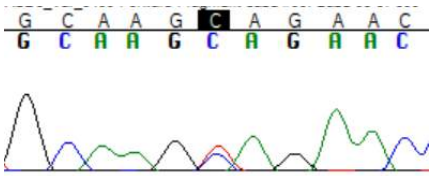
LZTR1_var_18:G>C



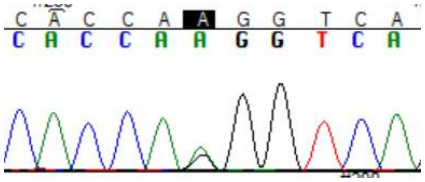
LZTR1_var_19:G>A



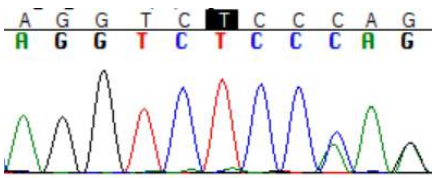
LZTR1_var_20:G>A



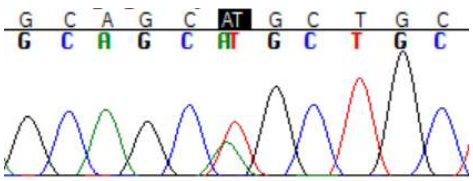
LZTR1_var_21:C>T



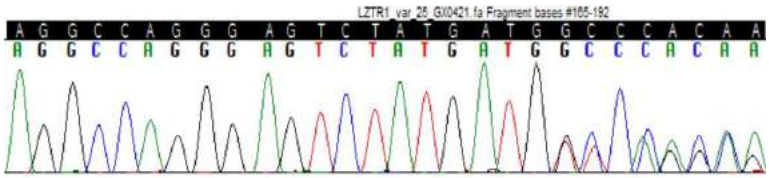
LZTR1_var_22:A>G



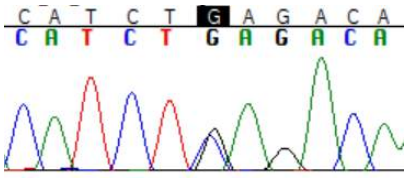
LZTR1_var_23:T>TC



LZTR1_var_24:T>A



LZTR1_var_25: A>AC



LZTR1_var_26:C>G

2.3 Pathogenic findings elucidation

We designed 80 smMIPs to cover all the annotated RefSeq coding exons as well as 5 bp of flanking intronic sequence for *ARVCF* and *LZTR1* gene, and we sequenced 11,731 cases with a primary diagnosis of ASD/ID/DD, for which exome sequencing had not previously been performed, from a large international collaboration between research and clinical investigators of 15 centres across the world. After QC, we identified and validated 38 novel LGD/MIS30s variants. Overall, the burden of these rare events retrieved in these two genes confirms their role in creating a sensitized genetic background for NDDs as predicted by our enrichment analysis.

Although we will discuss these data later in depth, so far it is worth pointing out that ARVCF belongs to the p120ctn family which includes 4 proteins encoded by four independent genes distinct from the more ubiquitously expressed α- and β-catenins. Evolutionarily, they are generated from an ancient 'δ-catenin-like' gene and divided into two major classes, one of which comprises p0071 and δ-catenin, while the other includes p120ctn and ARVCF (Carnahan RH et al., 2010). These genes are constitutively expressed in central neurons, and recently it has become clear that they have extensive functional roles in multiple aspects of neuronal morphology, synaptic structure, synaptic efficacy and molecular processes related to learning and memory (Seong E et al., 2015; Yuan L et al., 2017).

Conversely, even if *LZTR1* molecular mechanisms are still unclear, this gene is already ranked as a confident gene in the Simons Foundation Autism Research Initiative (SFARI) database. Indeed, an analysis of 2,377 families from the Simons Simplex Collection revealed its private LGD variants over-transmission in ASD probands (6 inherited CNVs/SNVs in probands compared to none in unaffected siblings; Krumm N et al., 2015).

CHAPTER 3

A phenotype-first approach to the 22q11.2 distal region

3.1 Does the DSM read the DNA?

The combined CNV and single-nucleotide variant approach performed in the previous chapter has been proven effective to highlight the neurodevelopmental role of several 22q11.2 genes. However, no genes harboured within the distal 22q11.2 region reached nominal significance to all the chosen enrichment criteria. This may be partially due to our strong selection analysis, which was also developed to detect the best candidates in the most powerful statistical assay without considering other known NDD pathogenetic molecular mechanisms such as the dosage-sensitivity.

Despite this evidence, it has also been shown that the five telomeric LCR22s, namely LCR22-D to H, are causally implicated in the recurrent distal 22q11.2 microdeletion (OMIM 611867) and the reciprocal microduplication-associated phenotypes (Shaikh TH et al., 2007, Tan T et al., 2011; Mikhail FM et al., 2014). These events are typically large (>0.5-1 Mb) and, are predominantly associated with neurocognitive disorders. To date, distal 22q11.2 rearrangements have been grouped together although these CNVs hold variable size and position due to different LCR22s-NAHR processing and, could be either partially overlapping or non-overlapping. According to Mikhail FM et al., 2014 these data suggest that they do not represent a single clinical entity rather than different nosological entries with some difference in both presenting features and, neurodevelopmental risks.

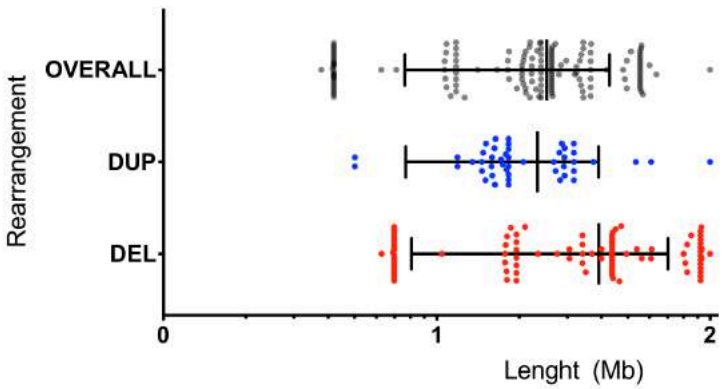
Moreover, although these rearrangements are often seen variably associated with a wide spectrum of congenital malformations, their large size prevents from claiming a single causative gene for a given disorder since multiple candidate may underlie the same region. Hence, because of the extreme 22q11.2 genomic locus heterogeneity, the pathogenicity of many distal CNVs observed in clinic remains unclear. Relatively few recurrences have been reported and, *de novo* overlapping CNVs are extremely rare requiring large surveys to achieve significance in case-control cohorts to determine an appreciable impact on human health (Cooper GM et al., 2011; Kaminsky EB et al. 2011).

Here, we aimed to identify novel NDD genes potentially sensitive to dosage imbalance within this critical chromosomal region by developing a phenotype-first approach. This model was designed to be specular to the one previously applied, allowing researchers to interface, anchor, and compare the phenotype features seen in a given genomic cohort to the genome. To this end, we built the uniquely available distal 22q11.2 CNVs "co-morbidity" map collecting the HPO (human project ontology) clinical information seen in patients carrying distal 22q11.2 rearrangements and subsequently depicting a common phenotype broken out *per* single LCRs22 and *per* both deletion and duplication events. When these results were integrated with published genomic sequencing data, we were able to identify new pathogenic events with both statistical significance and clinical relevance.

3.2 Building the distal 22q11.2 co-morbidity map

In order to build the distal 22q11.2 co-morbidity map we conducted a systematic review of the thematic scientific literature published from January 1980 to April 2018 using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Liberati A et al., 2018). A comprehensive literature search of the MEDLINE/PubMed, Cochrane Library, ClinicalTrial.gov, the Cumulative Index to Nursing and Allied Health Literature (CINHAL) databases was conducted. The search algorithm was based on a combination of the following terms: (22q11.2 distal deletion syndrome) OR (22q11.2 distal duplication syndrome) AND (Autism spectrum disorders OR Autism spectrum OR Autism OR Autistic Disorders OR Behavioural Disorders/Impairment) AND (Schizophrenia OR Early Onset Psychosis OR Early Onset of Psychosis Symptoms) AND (Developmental Delay OR Global Developmental Delay OR Intellectual Disability) AND (Congenital Malformation OR Congenital Anomalies OR Dysmorphisms). The last update of the search was performed at the beginning of May 2018, no language and/or study design limits were applied. As result, we found a pool of 59 studies on distal 22q11.2 rearrangements describing an overall amount of 148 patients whose genomic summary statistics is reported and depicted in Fig. 10. Each rearrangement, broken out *per* deletion or duplication, was then catalogued according to its: (i) reported genomic coordinates, (ii) D- to G- genomic conventional intervals defined between outer flanking LCRs22, and (iii) inheritance status (Supp. Tab. 11).

By the analysis of these data a wider rearrangement-size was detected in duplications (median ~1.37 Mb, mean ~1.63 Mb) than in deletions (median ~1.17 Mb, mean ~1.34 Mb) that represent more than half of the total rearrangements seen within the distal 22q11.2 region (del:dup ratio = 2.08:1).



LCR	DEL	DUP	OVERALL
D	51	3	53
D, E	17	1	18
D, E, F	11	0	11
D, E, F	0	3	3
D, E, F, G	2	1	3
E	14	2	15
E, F	3	1	4
E, F, G	1	16	17
F	1	0	1
F, G	2	21	23
OVERALL	100	48	148

	DEL	DUP	OVERALL
Number of values	100	48	148
Minimum (bp)	37060	431716	37060
25% Percentile (bp)	270068	1251202	682782
Median (bp)	1176307	1376220	1368197
75% Percentile (bp)	1855511	1941693	1919332
Maximum (bp)	3374785	5589899	5589899
Mean (bp)	1345337	1638114	1439025
Std. Deviation (bp)	1038456	809143	977900
Std. Error of Mean (bp)	102822	116789	79845
Lower 95% CI of mean (bp)	1141364	1403163	1281250
Upper 95% CI of mean (bp)	1549309	1873064	1596800

Figure 10: On the top (left): ideogrammatic representation of the 22q11.2 rearrangements length (Mb) distribution seen in 148 patients broken out *per* overall (grey dots), deletion (red dots) and duplication (blue dots) events; X axis values plotted in Log 10 scale after data normalization, statistical analysis was performed in these sets using the GraphPad Software, Inc. On the top (right): LCRs22 overlapping and/or non-overlapping rearrangements' count. Below (in table): 22q11.2 distal rearrangements' summary statistic.

Moreover, while most events are harboured between the centromeric LCRs22 and decreasing progressively through the region (R squared: 0.9837), a different rearrangements distribution pattern arose for both deletion and duplication events (Fig. 11). Indeed, 22q11.2 distal deletions predominantly occurred between the D- and E-LCRs22 (R squared: 0.9746) contrary to the reciprocal duplications whose frequency peaks in a more telomeric position, i.e. between the F- and G-LCRs22 (R squared: 0.8191).

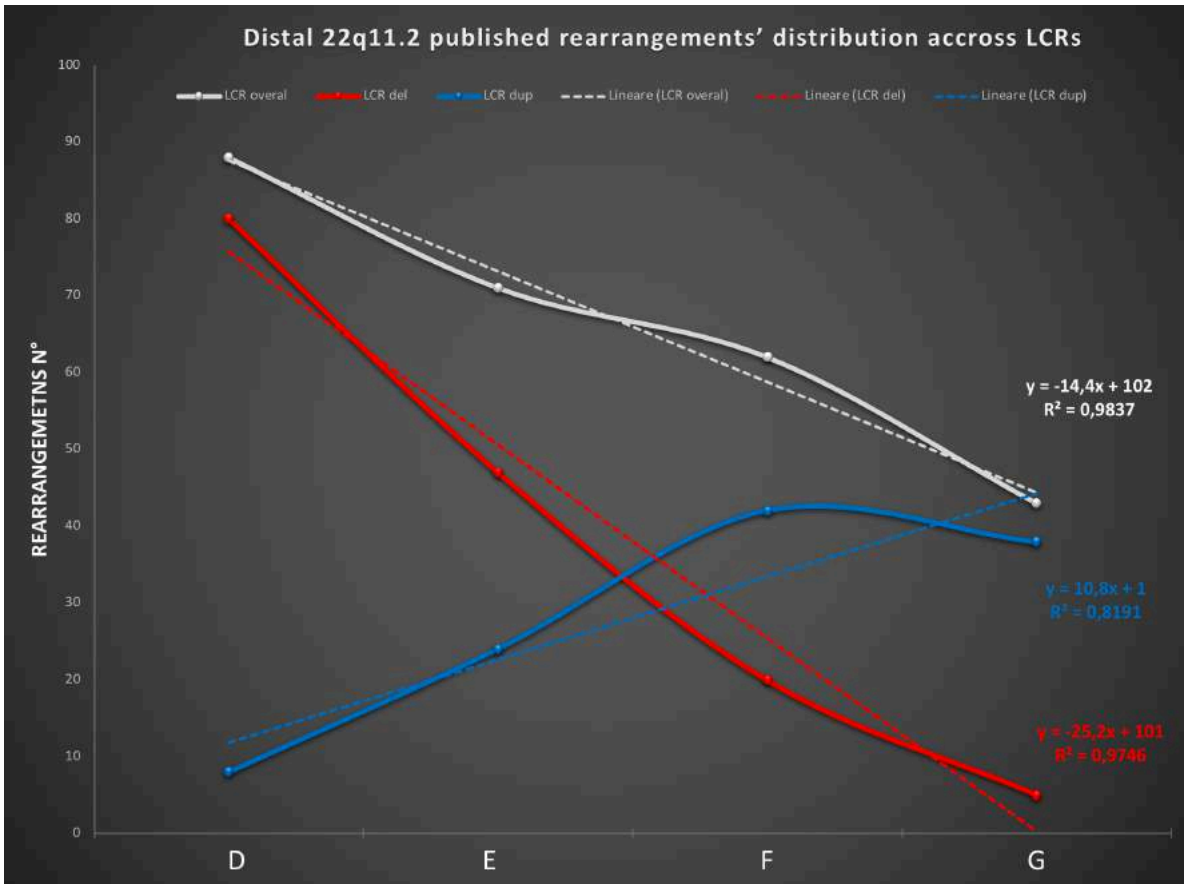
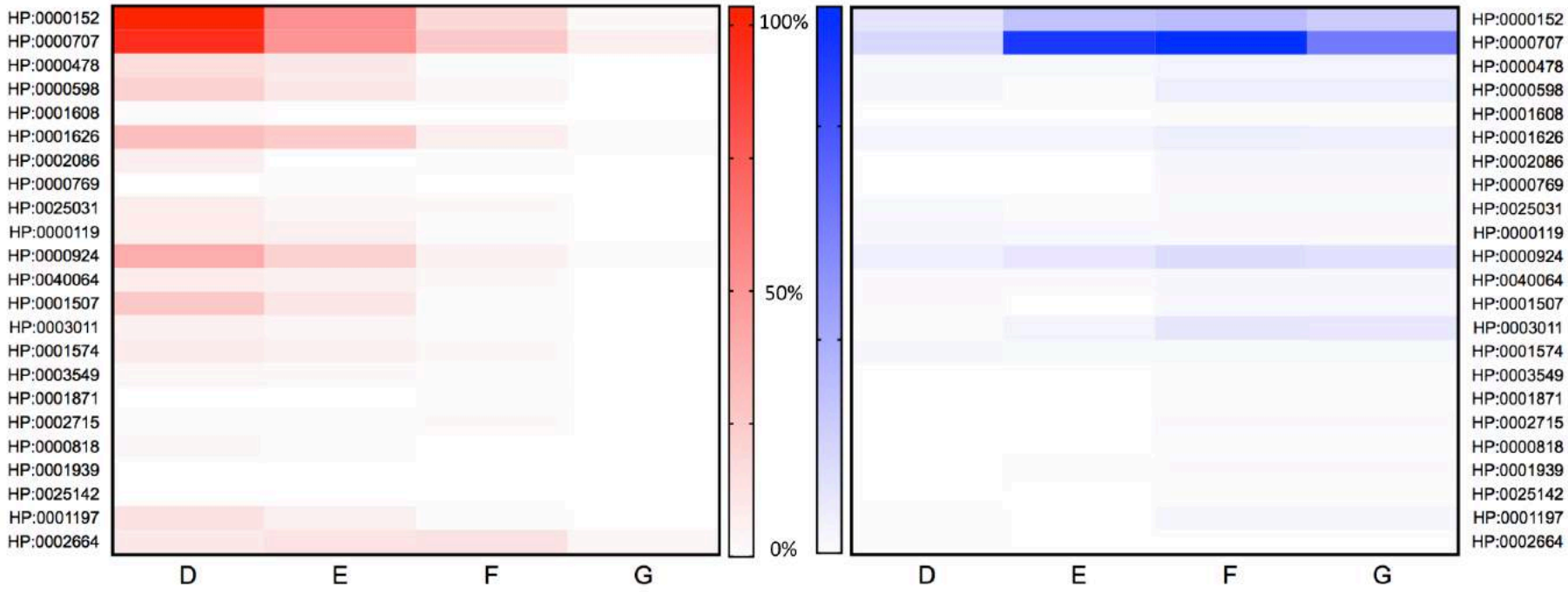


Figure 11: Overall 22q11.2 distal rearrangements distribution (grey lines) across LCRs22 and broken out *per* deletion (red lines) and duplication (blue lines) events. On Y axis, we show the number of events retrieved. Dashed lines depict linear regression analysis results of the individual categories examined.

Hence, aimed to highlight the clinical picture driven by each distal event, we first collected any phenotypic features and/or clinical information from the 148 available patient reports. These elements were subsequently translated into the corresponding specific HPO codes (Supp. Table 12), and then assigned to the belonging HPO macro-categories (HPO phenotypic abnormalities' major subclasses or HPO nodes "0") by using the Charité HPO browser (Köhler S et al., 2017). Finally, each HPO node "0" event was matched to the corresponding genomic-driven rearrangement allowing us: (i) to build a co-morbidity map of the patients' features and, (ii) to indirectly localize the more pathological intervals represented in this cohort (Fig. 12, Supp. Table 13). Indeed, by the statistical analysis of these data, we found that the highest features' signatures are respectively clustered at the distal 22q11.2 D interval for deletion (55% of all the features retrieved, which represent an enrichment factor of 2.21 times higher than expected by a random chance assignment), and between the E to F intervals for duplication events (62.5% of all the features retrieved, enrichment factor of 1.25 times higher than expected) both mainly characterizing a neurodevelopmental disorder.

Figure 12: Heatmap representation of the 22q11.2 distal co-morbidity map, built through HPO phenotypic abnormalities' count retrieved from our literature review and, depicted *per* HPO node "0" major subclasses (Y axis) matched to the previously described LCRs22 conventional genomic intervals (X, axis). Vertical coloured bars represent the feature frequency scale broken out *per* deletion (red) and duplication (blue) events. In picture HPO legend as follow: Abnormality of head or neck, HP:0000152; Abnormality of the nervous system, HP:0000707; Abnormality of the eye, HP:0000478; Abnormality of the ear, HP:0000598; Abnormality of the voice, HP:0001608; Abnormality of the cardiovascular system, HP:0001626; Abnormality of the respiratory system, HP:0002086; Abnormality of the breast, HP:0000769; Abnormality of the digestive system, HP:0025031; Abnormality of the genitourinary system, HP:0000119; Abnormality of the skeletal system, HP:0000924; Abnormality of limbs, HP:0040064; Growth abnormality, HP:0001507; Abnormality of the musculature, HP:0003011; Abnormality of the integument, HP:0001574; Abnormality of connective tissue, HP:0003549; Abnormality of blood and blood-forming tissues, HP:0001871; Abnormality of the immune system, HP:0002715; Abnormality of the endocrine system, HP:0000818; Abnormality of metabolism/homeostasis, HP:0001939; Constitutional symptom, HP:0025142; Abnormality of prenatal development or birth, HP:0001197; Neoplasm, HP:0002664.



3.3 NDD patients' cohort genomic screening by CGH Array

Since the highlighted intervals were retrieved leading to different neurodevelopmental disorders variably associated with mild degree of dysmorphisms and congenital abnormalities, and mainly determined by a D deletion and, E or F region duplication, we interbreed these results with the previously described CNV morbidity map (Coe BP et al., 2014) to narrow their respective pathogenic driver regions. As result, two small regions of interest reached the highest nominal significance through the distal 22q11.2 overall loci rearrangements signatures' analysis, exactly arising where predicted by our phenotypic-first approach (Fig. 13).

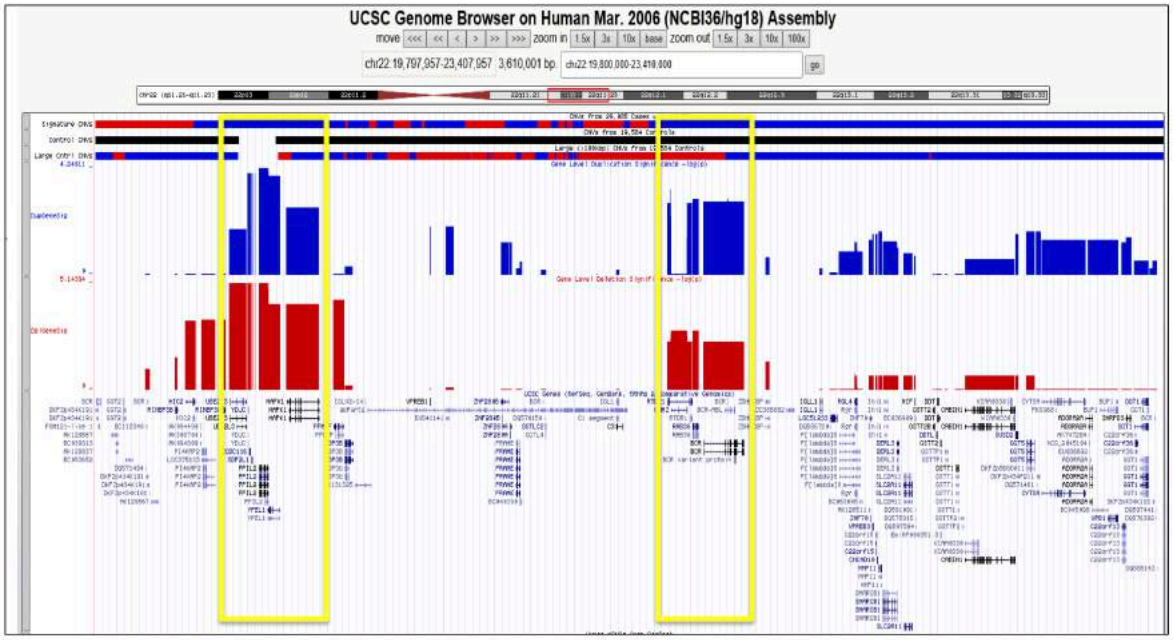


Figure 13: 22q11.2 distal CNV morbidity map representation depicted by using the UCSC Genome Browser and a customized truck hub reported in Coe BP et al., 2014 courtesy of the Eichler Lab, Department of Genome Sciences at University of Washington in Seattle USA. Above: each line represents packed genomic tracks of those events seen in the 29,085 ID/DD cohort of patients (first row), and 19,584 controls respectively broken out *per* overall (second row), and large (>100Kb, third row) events. Rearrangements coordinates plotted according to the Human Mar. 2006 (NCBI36/hg18) Assembly. Below: Fisher exact test logarithmic representation of the 22q11.2 distal genomic loci rearrangement's signature (Suppl. Tab. 9) plotted breaking out deletion (red rectangles) and duplication (blue rectangles) values. In figure yellow rectangles highlight our regions of interest described in text and those NCBI RefSeq genes harboured within.

As a matter of fact, we identified:

1) in the D interval, a ~300 Kb genomic frame ranging from 20,251,957 to 20,551,970 (NCBI36/hg18 coordinates) at the 22q11.21 sub-band which harbours the *UBE2L3, YDJC, CCDC116, SDF2L1, PPIL2, YPEL1* and *MAPK1* genes and, whose loci

rearrangements are seen in 51 patients versus 10 in controls (One-tailed Fischer *p-value*: 5.418e-05, Odd Ratio: 3.438);
2) in the E interval, a ~256 Kb genomic frame ranging from 21,731,593 to 21,990,224 (NCBI36/hg18 coordinates) at the 22q11.23 sub-band which harbours the *RTDR1, GNAZ, RAB36* and *BCR* genes and, whose loci rearrangements are seen in 51 patients versus 8 controls (One-tailed Fischer *p-value*: 7.648e-06, Odd Ratio: 4.298).

Thus, according to our model and assuming that the neurological impairment observed in our co-morbidity map could have been determined by an altered dosage effect of one or more genes harboured within the above-mentioned candidate small frames of interest, we performed a new NDD patients' cohort enrolment to find out the best matching genomic events, and to further investigate how the predicted imbalance may affect the behaviour of neuronal circuits. To this end, from January 2016 to January 2018, we screened a large cohort of NDD children referred for genetic evaluation to both the Medical Genetics Unit of the Department of Experimental Medicine, Sapienza University of Rome, Italy and, to the Regional Referral Centre for Rare Genetic and Chromosomal Diseases, AOOR "Villa Sofia-Cervello" Hospital of Palermo, Italy as part of a joint research project. As result, although we did not retrieve overlapping deletion events in the distal D genomic interval, CGH analyses performed on all recruited patients allowed us to identify two male children (from this point onwards referred as "patient I" and "patient II") carrying a ~0.5 Mb microduplications at the 22q11.22q11.23 sub-band that were seen fully incorporating our predicted "telomeric" frame of interest (Fig. 14).

However, because of these CNVs were in both cases inherited and actually classified as variations of uncertain significance, our patients were subjected to an extensive diagnostic workup that included targeted gene tests sequencing (*FMR1* and *MECP2*), and the inborn errors of metabolism (IEMs) recommended screening tier for treatable DD/ID (Suppl. Tab. 14). These evaluations did not disclose the presence of any further recognizable etiologic cause of ASD, DD/ID and/or seizures/epilepsies according to the current international recommendations (van Karnebeek CD et al. 2014; Moeschler JB et al. 2014). Patient I and patient II medical history were characterised by the record of delayed neurodevelopment milestones acquisition during pre-school age. Moreover,

patient I was diagnosed at age of 6 years with moderate intellectual disability (DSM-5, full IQ on WISC IV-R: 50) while patient II at age of 4 years showed global development delay (Griffiths/Cleveland scale) in addition to an ASD diagnosis (DSM-5, score 8 on ADOS-2 assessment).

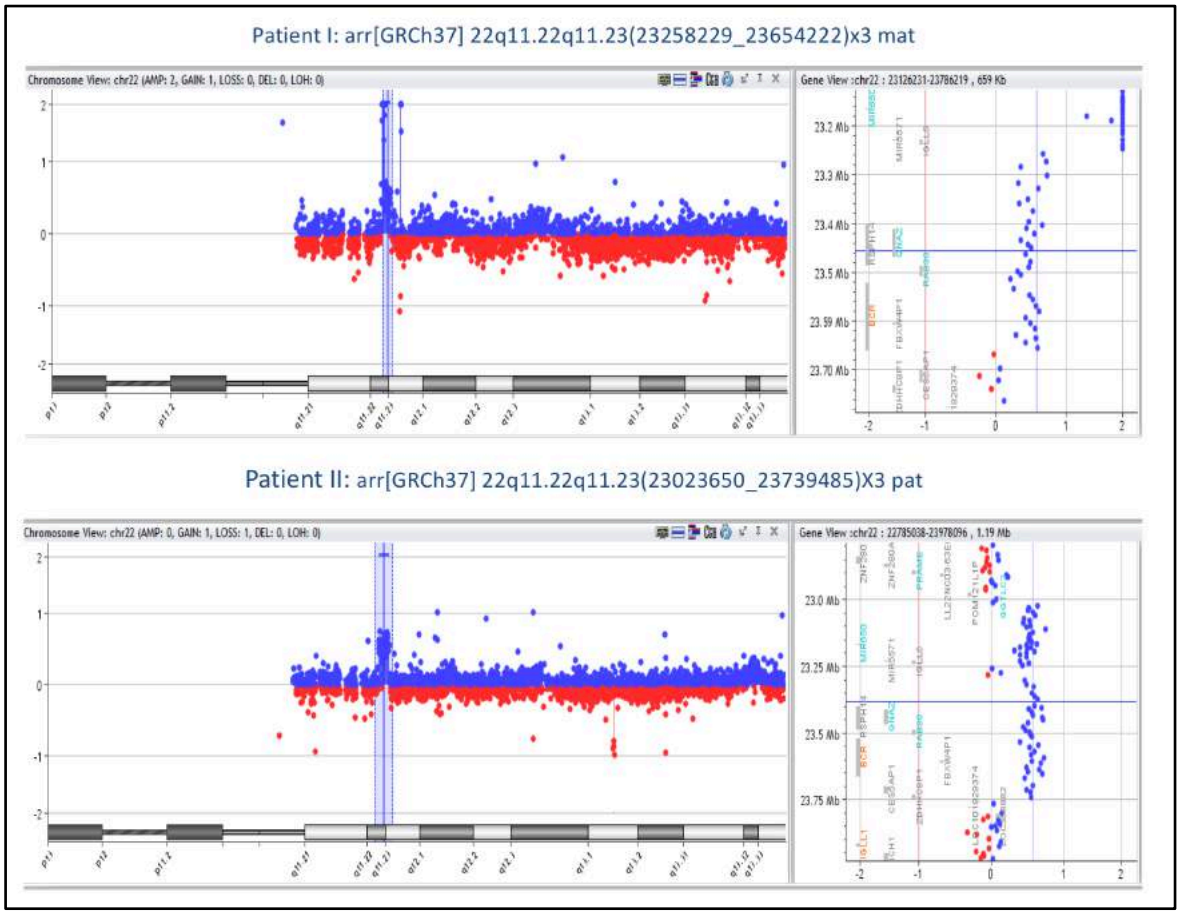


Figure 14: Patient I (on the top) and II (down) CGH Array analyses results and, 22q11.2q11.23 CNVs' genomic coordinates according to the Human Feb. 2009 (GRCh37/hg19) assembly.

Our patients' clinical evaluation included visual and hearing tests, echocardiogram, abdominal ultrasound, skeletal survey and X-ray bone age studies which were normal as well as their anthropometric parameters. Their further neurophenotype characterisation through brain MRI imaging and electrophysiological examinations showed normal myelinisation and CSF volume without structural anomalies but, noteworthy, unmasked the same concurrently items of frontal spike wave and multiple spike discharges on sleep-deprived EEGs (Fig. 15) as ones retrieved in other four previously published carrying patients, whose duplications have been detected up to 0.8 Mb at the same 22q11.22q11.23 sub-band (Piccione M et al., 2011, Vecchio D et al., 2016). Furthermore, at the molecular level, their array-CGH comparison showed a 0.5

Mb small overlapping region (SRO) ranging from ~23,100 to ~23,600 Mb (GRCh37/hg19 assembly) harbouring 1 microRNA (*hsa-miR-650,* OMIM:615379) and 5 genes (*IGLL5,* OMIM:146770; *RTDR1*, OMIM:605663; *GNAZ,* OMIM:139160; *RAB36*, OMIM: 605662; and *BCR,* OMIM:151410; Fig. 16), prompting us to further investigate the causative pathogenic mechanism trough their expression analyses.

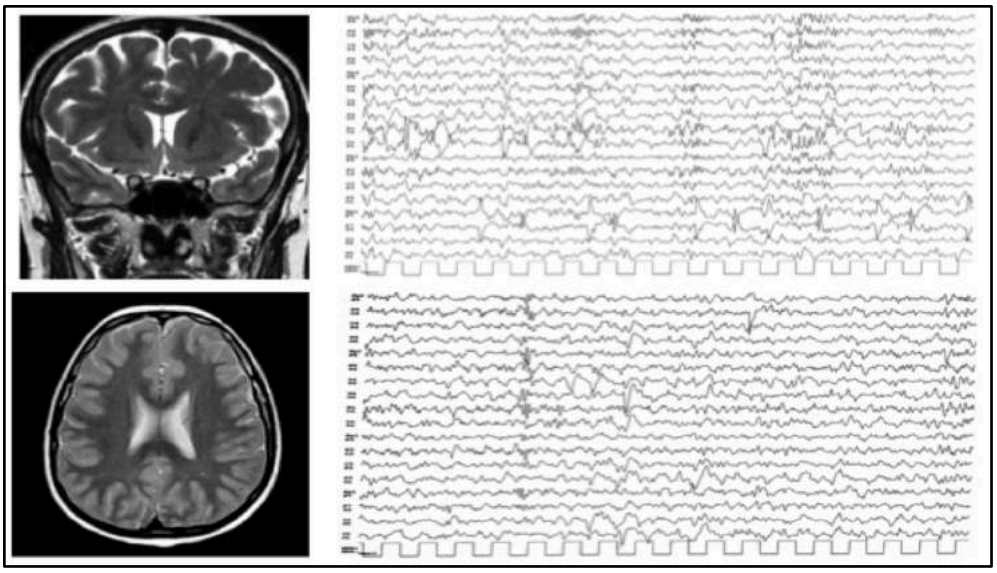


Figure 15: MRI and sleep-deprived patients' EEGs imaging showing no brain structural anomalies and same items of frontal spike wave and multiple spike discharges.

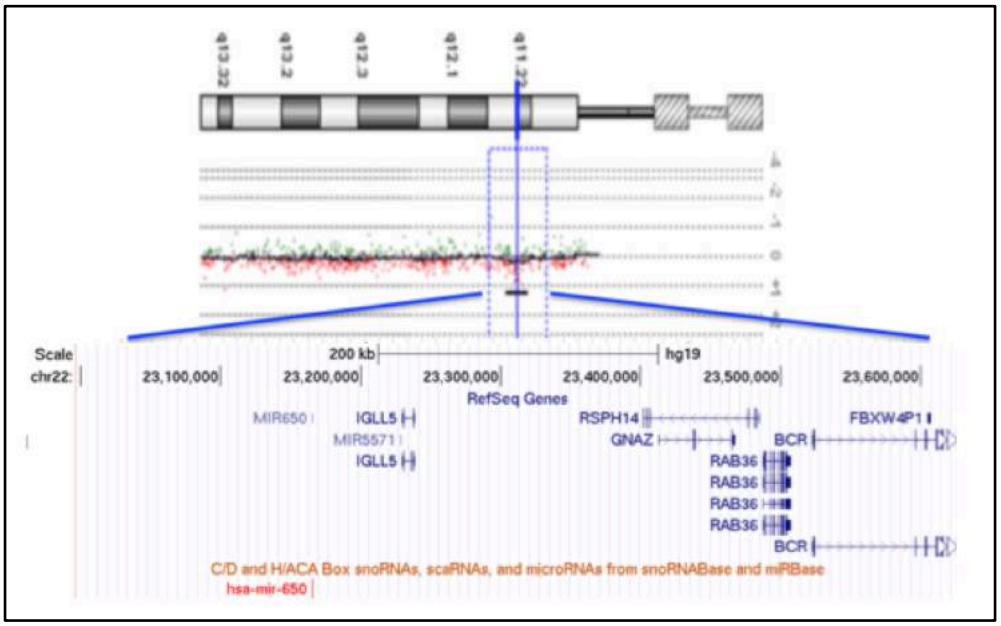


Figure 16: Above: First reported 0.6 Mb microduplication at the 22q11.22q11.23 chromosome Array-CGH analysis result (courtesy of Piccione et al., 2011). Below: ideogrammatic representation of the loci harboured within the identified SRO at the 22q11.23 chromosome depicting by using the UCSC Genome Browser and the Human. Feb. 2009 - GRCh37/hg19 Assembly.

3.3.1 SRO genes expression analysis

The phenotypic characterization of patients carrying distal 22q11.22q11.23 microduplications discloses a common clinical picture mainly characterized by DD/ID, ASD and peculiar electroencephalographic elements. In this study the enrolment of two new patients, with same clinical features and overlapping rearrangements, suggested once again the neurodevelopmental role of this genomic region and, on molecular level, allowed us to highlight a SRO that we assumed critical for the observed phenotypes. Thus, a tissue-specific expression analysis of genes and microRNAs mapped within was first performed according to available gene/protein expression databases exhibiting a brain expression for 4 of them (*RTDR1, GNAZ, RAB36* and *BCR*; Tab. 6). In this view, although a pathogenic role for *RTDR1*, *GNAZ* and *BCR* genes cannot be excluded, the *RAB36* gene, which holds a documented function in neurons, emerged as the best candidate.

Indeed, RAB36 protein has been recently described being involved in vesicular trafficking and, its impairment has been proposed interfering in neurotransmitters processing, neural secretion and/or maturation processes (Kobayashi H et al., 2014) resembling several yet characterized impaired pathways described in other DD/ID RAB-related disorders. Hence, as following step and to ascertain its effector role in the observed pathogenetic mechanism, we decided to assess and to compare the brain-encoded SRO genes expression in our patients, in their carrying parents and in a negative control pool of individuals.

Table 6: Expression of genes/microRNAs mapping within the SRO at the 22q11.23 sub-band in brain according to several gene/protein expression databases (in table: +, both mRNA and/or protein presence; -, both mRNA and/or protein absence; NR, data not recorded).

	TSRI BioGPS Database (biogps.org)	The Human Protein Atlas (proteinatlas.org)	Allen Brain Atlas (human.brain-map.org)	miRmine Human miRNA Expression Database (guanlab.ccmb.med.umich.edu)
IGLL5	-	-	-	NR
RTDR1	+	+	+	NR
GNAZ	+	+	+	NR
RAB36	+	+	+	NR
BCR	+	+	+	NR
3mIR-650	NR	NR	NR	-

3.3.2 Experimental procedure

Expression analyses of the above mentioned brain-encoded genes (*RAB36, IGLL5, RTDR1, GNAZ* and *BCR*) mapping in the highlighted SRO at the 22q11.23 chromosome were first verified on a human cDNA brain library extracted from brain tissue of control subjects, and then performed on blood samples withdrawn from patients, their carrying parents and a pool of age-matched negative controls through RT-qPCR assays.

RNA isolation and cDNA synthesis

Total RNA was extracted from peripheral blood cells (PBCs) using the QIAamp RNA Blood Mini Kit (Qiagen) following manufacturer's instructions. RNA concentrations and quality were verified by spectrophotometry (optical density (OD) at 260 nm), whereas the RNA integrity was checked using a 1.5% agarose gel. The RNA was stored at −80 °C for future use. The extracted RNA (1 μg) was treated with RNA qualified 1 (RQ1) RNase-Free DNase (Promega, Madison, WI, USA) to remove any residual genomic DNA contamination, and the DNase was inactivated by adding 25 mM EDTA. First-strand cDNA was synthesised from 100 ng DNase-treated total RNA samples using oligo(dT)18 and Superscript III (Invitrogen Corporation, Carlsbad, CA, USA) following the manufacturer's instructions. The cDNA mixture was stored at −20 °C. The absence of any residual genomic DNA contamination in each cDNA preparation was checked performing minus-reverse transcriptase ("-RT") control in RT-PCR experiments with the different primer sets.

Relative quantification using real-time quantitative polymerase chain reaction

RT-qPCR was performed using the ABIPRISM 7500 System (Applied Biosystems, Forster City, USA) with Power Sybr Green as detection chemistry (Applied Biosystems, Forster City, USA). The qPCRs were carried out on blood cells cDNA and a human brain cDNA preparation Human MTC™ Panel I (Takara Bio USA, Inc.) with primers reported in Supp. Tab. 15. The housekeeping gene 18S rRNA, and HPRT1 were chosen as reference genes. Their expression stability was evaluated using the GeNorm software and they were selected as internal controls in our experiments. A GeNorm normalization factor was calculated based on the expression level of 18S rRNA, and HPRT1 and used to quantify the expression levels of the target genes. Quantitative

real-time PCR was conducted according to the manufacturer's recommended procedures, and every reaction was repeated in triplicate. The amplification conditions were the following: initial denaturation at 95 °C for 10 min and 40 cycles of 95 °C for 30 s and 60 °C for 50 s, followed by a melting curve from 60 to 95 °C. The qPCRs were performed in 20µl reactions containing 0.5 µM of each primer set and 1× Power Sybr Green (Applied Biosystems, Forster City, USA). Amplicons were detected by agarose gel analysis after each PCR to confirm the amplification of the specific gene. All data represented relative mRNA expressed as the mean ± S.D. (n=3). Statistical data analysis was performed using the Prism software package, version 6.

Legal authorization

Samples of peripheral blood from patients, parents and controls were achieved under patients and/or parents' consent according to the Helsinki Doctrine for Human Experimentation and the Italian Law: "Codex on the protection of personal data" (DL n°196, 30/06/2003). The study was approved by the Ethic Committee *Palermo2* (Comitato Etico *Palermo2* AOR VILLA SOFIA CERVELLO – ARNAS CIVICO – ASP TP – ASP AG at the AOR Villa Sofia-Cervello Hospital, Palermo, Italy), number of the folder: 3994/2017/CRR. Patients' parents signed an informed consent before enrolment and genetic testing.

3.3.3 Results

The expression analyses of candidate genes performed on a human cDNA brain library confirmed the *in silico* data retrieved from our previous gene expression database searching. All genes were detected encoded in the human brain under physiological conditions, with a wide range of expression levels (Fig. 17). In order to evaluate if the 22q11.22q11.23 microduplication could affect the expression level of the candidate genes, expression assays were also performed on patients' blood. As a first step, the IGLL5, RTDR1, GNAZ, RAB36 and BCR transcripts were quantified in PBCs obtained from two healthy individuals as controls (control I and II) to evaluate potential fluctuation of transcript level in physiological conditions. Analysis of the transcript levels for each gene resulted similar among controls suggesting a stable expression pattern for all genes (Fig. 18). These results allowed to conclude that

expression analyses of candidate genes in PBCs could represents a reliable analysis for gene expression level comparisons.

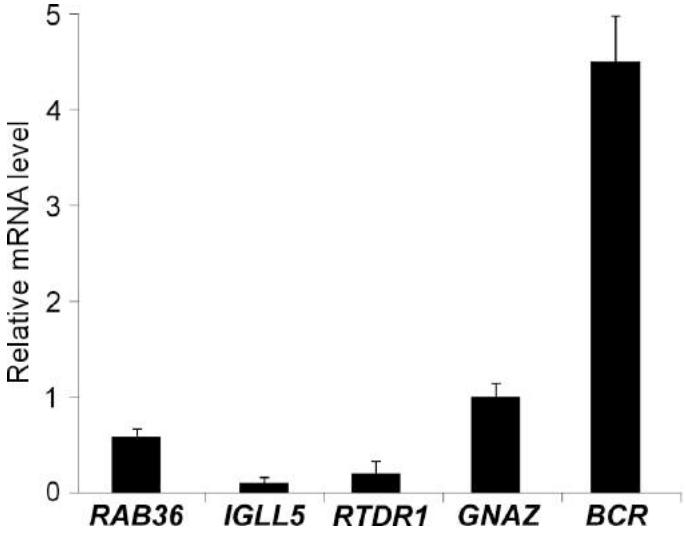


Figure 17: Expression levels of *IGLL5, RTDR1, GNAZ, RAB36* and *BCR* genes in brain (black pales) using *HPRT1* and *18S* as reference genes. The qPCR-derived expression of genes was normalised to that of *GNAZ*. The results are represented as means ± SD (n = 3).

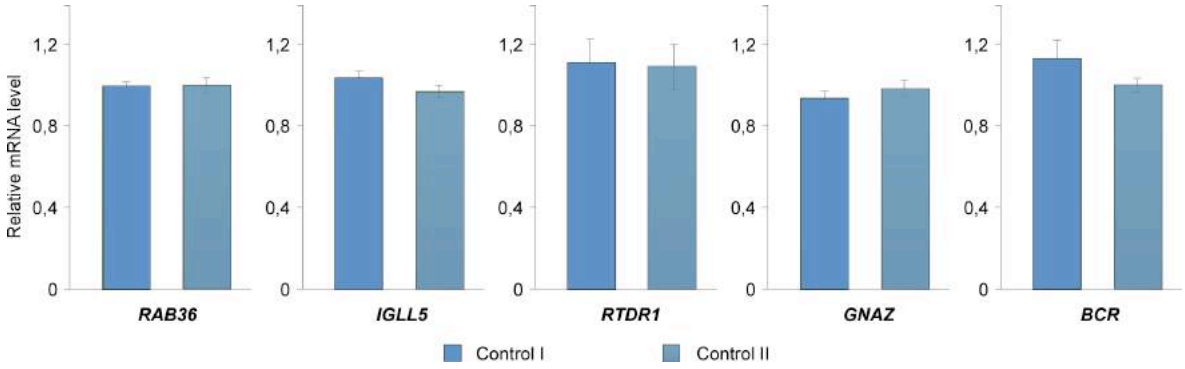


Figure 18: qRT-PCR evaluation reveals stable levels of the IGLL5, RTDR1, GNAZ, RAB36 an BCR mRNA in PBCs obtained from two male controls samples (blue and green pales).

Since stable mRNA expression of the different genes were observed in controls, cDNAs obtained from controls' PBC were pooled and used for comparative evaluation to investigate the occurrence of dosage dependent expression in patients with microduplications of 22q11.23. The mRNA levels of the candidate genes were then analysed in PBCs derived from two patients, relative parents and two healthy controls.

As result, parents carrying the microduplication of the critical region showed a slight non-significant overexpression for all genes tested. Differently, the two patients disclosed a significant increased expression levels only for the RAB36 gene (Fig. 19), both compared to controls and to respective parents ($p < 0.001$, FDR 0.01, multiple t tests). These data confirm that RAB36 may play a key role in the pathogenesis of the observed patients' neurophenotypes acting in a dosage-dependent manner.

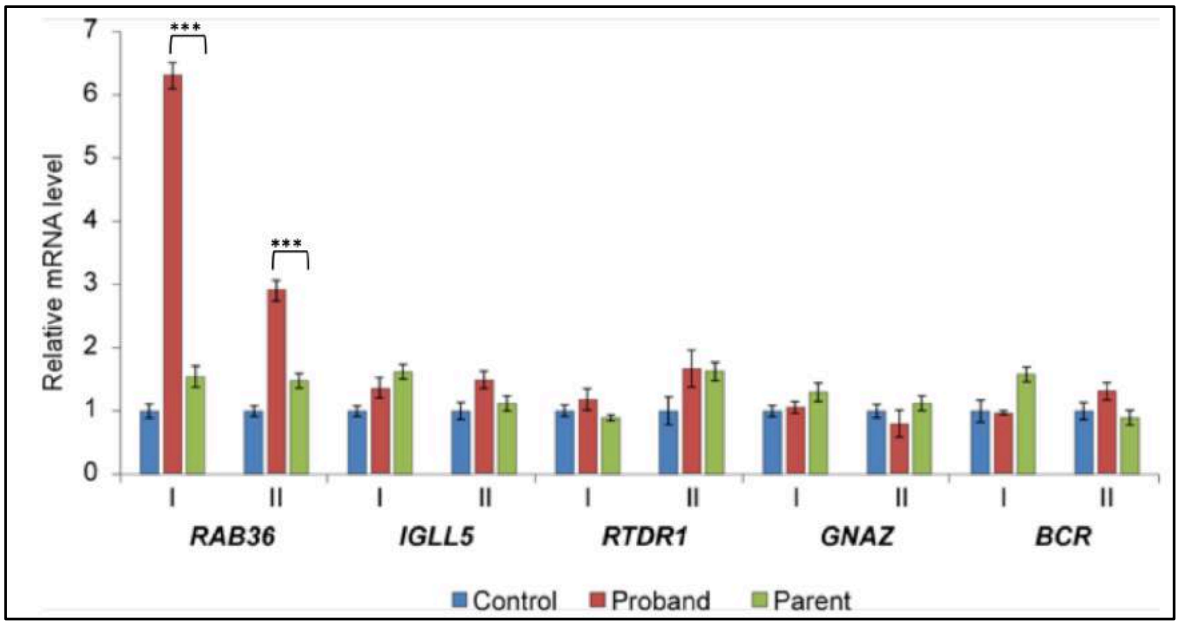


Figure 19: Expression analyses of selected genes in two probands (I and II, red bars) and two related parents (pale green bar) measured by qRT-PCR relative to two male controls (pale blue bars). Data are the mean ± SD of tree independent experiments normalized to the housekeeping genes HPRT1 and 18S; ***p < 0.001, multiple t tests, FDR 0.01, n = 3 technical replicates *per* group.

3.4 Candidate gene overexpression analysis

Despite substantial progress in understanding the role played by RAB GTPases and their effector proteins on cognitive disorders' pathological mechanisms, to date there is still the need of a more systematic characterization of the brain-expressed RAB GTPases to deeply understand their driven function during the CNS development and maturation (Mignogna ML et al., 2015). This is also influenced by the lack of pre-clinical models that could offer the possibility to evaluate and to validate an experimental approach. In this view, small mammals' neuronal cell lines could be of potential help to mimic the human condition and to establish an appropriate model that can be used for studies on human brain development disorders like those

retrieved in our distal 22q11.22q11.23 rearranged patients. In this view, and to ascertain the *RAB36* neurodevelopmental role, we decided to further investigate the Rab36 overexpression effect on neuronal differentiation by studying a customized *in vitro* mouse neuronal specimen. This experiment was carried out at the Molecular Genetics of Intellectual Disabilities Unit (D'Adamo Laboratory), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy as part of a joint research project.

3.4.1 Experimental plan

Experiments were carried out on primary E18 mouse hippocampal neuronal cultures, transfected after 2 days in vitro (DIV) with a plasmid expressing Rab36 tagged with the green fluorescent protein (GFP), from at least three independent preparations. At 7 DIV we quantified the level of GFP-Rab36 expression by Western blot using an anti-RAB36 antibody (ABCAM, ab191531, data not shown) and then, to evaluate the degree of neuronal differentiation, E18 mouse primary hippocampal neuronal cultures overexpressing GFP-RAB36 were fixed, stained for phalloidin and subjected to a Sholl analysis in order to characterize the neurons morphological features assessing their: mean diameter, dendrites or axons length and, number of their extensions (Sholl DA et al.) in comparison to a control negative (wild type) culture.

3.4.2 Materials and Methods

Plasmid generation

pFLAG-RAB36 was created amplifying the full-length mouse *Rab36* coding sequence using specific primers (RAB36 For: 5'-GCGAAGCTTATGAGGTCCTCTTGGACCCCT-3'; RAB36 Rev: 5'-GCCGGATCCTTAACAGCAGCCTAGGCCGGG-3'). The PCR product was then cloned into HindIII and BamHI sites in pFLAG CMV2 plasmid in frame of FLAG tag.

Animals

C57Bl/6N mice were obtained from Charles River, Italy and euthanized in accordance with "Institutional Animal Care and Use Committee San Raffaele (IACUC)" at San

Raffaele Scientific Institute, Milan, Italy, approved by the Italian National Ministry of Health, IACUC ID 470 and in accordance with the guidelines established by the European Community Council Directive of 24 November 1986 on the use of animals in research (86/609/EEC). All animals were maintained on a 12 h light/darkness cycle at 22–25°C. Food pellets and water were available *ab libitum*, unless stated otherwise. Wild Type (WT) mice were used.

E18 hippocampal neuronal cultures

Primary neuronal cultures were prepared from the hippocampi of E18 embryos from C57BL/6N mice. After 10 minutes of incubation with 0.25% trypsin (Sigma T1005-1G) in Hanks' balanced salt solution (HBSS; Gibco 14170) at 37°C, hippocampi were washed three times with HBSS to remove trypsin and then mechanically dissociated. Neurons were counted, and 180,000 cells/coverslip were plated on poly-L-lysine (Sigma-Aldrich P2636; 0.1 mg/ml)-treated glass coverslips (24 mm diameter). Cells were plated in NM5 [Neurobasal (Life technologies 21103) supplemented with 5% Foetal Bovine Serum (FBS, Life Technologies 16000), 2 mM glutamine (Life Technologies 25030) and 100ug/ml Peniciline/Streptomycine (Life technologies 15140) and 2% B27 (Gibco 17504)] and incubated for 5 hours at 37°C in a 5% CO_2 humidified atmosphere to allow adhesion to the substrate. 5 hours after plating, medium was changed with NM0 [Neurobasal (Life technologies 21103) supplemented with 2 mM glutamine (Life Technologies 25030) and 100ug/ml Peniciline/Streptomycine (Life technologies 15140) and 2% B27 (Gibco 17504)].

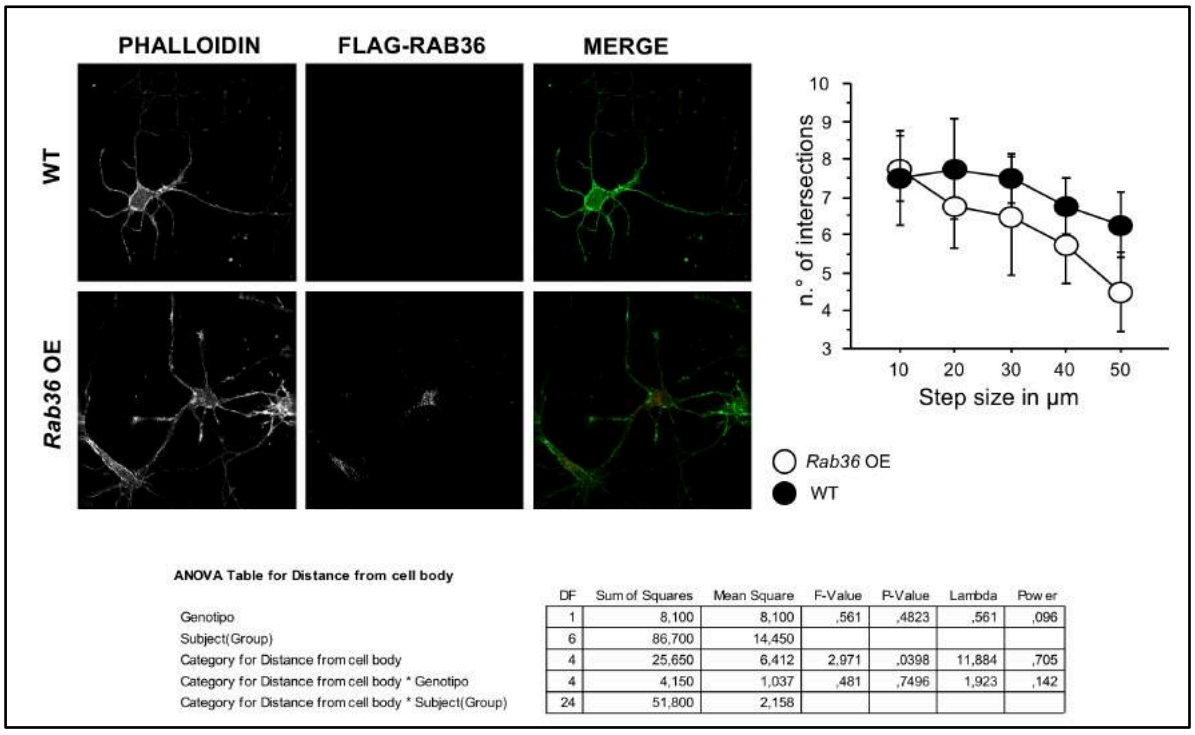
Transfection and immunofluorescence

Mouse hippocampal neurons were transfected at DIV 2 with pFLAG-RAB36 using Lipofectamine 2000 (Life-Technologies 11668) following manufacturer instructions. Standard immunofluorescence experiments were carried out as previously described (Giannandrea M et al., 2010). Briefly, cells were fixed at DIV 7 for 15 minutes with 4% paraformaldehyde (Sigma-Aldrich P6148), 4% sucrose (Sigma-Aldrich S5016) in 120 mM sodium phosphate buffer, pH 7.4. Coverslips were rinsed 3 times with phosphate-buffered saline (PBS) and then incubated 3 hours into a humidified chamber with the primary anti-flag monoclonal antibody (Sigma-Aldrich F1804) appropriately diluted in goat serum dilution buffer (GSDB; 15% goat serum, 450 mM NaCl, 0.3% Triton X-100,

20 mM sodium phosphate buffer, pH 7.4). Coverslips were washed 3 times within 30 minutes with high salt buffer (HS; 500 mM NaCl, 20 mM sodium phosphate buffer, pH 7.4) and then incubated with Alexa Fluor 568 goat anti-mouse secondary antibodies (Molecular Probes, Invitrogen A11004) for 90 minutes at room temperature. Together with the secondary antibody, Alexa Fluor 488 Phalloidin (Molecular probes, invitrogen A12379) was used to stain neuronal structure. After three washes with HS over 30 minutes and one wash with 5 mM sodium phosphate buffer, pH 7.4, coverslips were mounted with Vectashield (Vectalab).

3.4.3 Results

As result, the GFP-Rab36 overexpressed primary E18 mouse hippocampal neuronal cultures showed a morphological maturation impairment due to lower dendrite branching (Fig. 20) in comparison to the WT culture obtained under same condition procedure. This data confirms that Rab36 may play a key role in NDD pathogenesis acting in a dosage-dependent manner.



ANOVA Table for Distance from cell body

	DF	Sum of Squares	Mean Square	F-Value	P-Value	Lambda	Power
Genotipo	1	8,100	8,100	,561	,4823	,561	,096
Subject(Group)	6	86,700	14,450				
Category for Distance from cell body	4	25,650	6,412	2,971	,0398	11,884	,705
Category for Distance from cell body * Genotipo	4	4,150	1,037	,481	,7496	1,923	,142
Category for Distance from cell body * Subject(Group)	24	51,800	2,158				

Figure 20: DIV 7 wild type (WT, black docs) and Rab36 overexpressed (Rab36OE, white docs) E18 hippocampal neuronal cultures analysis' results (detailed in text). Pictures of immunostained hippocampal neurons were captured with the same exposure conditions using Delta Vision microscope (Applied Precision) equipped with a 60X objective. Z-space slices (0.3-0.4 µm) were deconvolved and flattened by maximum projection. In table: data expressed as mean ± standard error mean. Statistical significance was assessed using ANOVA analysis by using Statview 5.0 (SAS Institute, Cary, NC, USA, www.statview.com).

DISCUSSION

Deep genotyping of patients carrying different 22q11.2 rearrangements was successfully demonstrated a valid approach to uncover the genetic basis of several 22q11.2DS and 22q11.2DupS co-morbidities and/or previously categorized idiopathic diseases spanning from congenital malformations to specific psychiatric disorders (Lopez-Rivera E et al. 2017, Greene C et al. 2017). In this research project, we applied a genotype-based and a phenotype-first approach to further define, hence to experimentally confirm, the pivotal neurodevelopmental role of this region. When these models' results have been integrated with previously published genomic sequencing data, we were able to identify new pathogenic events with both statistical significance and clinical relevance adding new insights on elucidation of 22q11.2DS/22q11.2DupS neurobehavioral outcomes as well as of NDD neuroscience physiopathology.

Through our genotype-first approach we identified 8 genes (*ARVCF, CLDN5, DGCR8, RANBP1, TRMT2A, MED15, AIFM3* and, *LZTR1*) harboured within the 22q11.2 region that, according to our enrichment analysis criteria, can exert a downstream functional convergence upon those circuits involved in neurodevelopmental physiological and pathophysiological behaviours. Moreover, to functionally ascertain that this genes' set could be thought as NDD high-risk or high-impact candidate genes, we further explored their expression in fetal frontal cortex. Indeed, this area has been recently investigated with respect to cognition, neurodevelopmental and psychiatric disorders since several neurobehavioral causative genes have been described early activated in pathways critical to its neurogenesis (Gulsuner S et al., 2014, Jaffe AE et al. 2018). As result, all our intersected genes set was retrieved consistently expressed from early fetal frontal cortical elements (8 post-conception weeks) and throughout its development, with expression declining after the 24 post-conception weeks, and rising again in childhood mostly peaking between 2 and 3 years-old that coincides with the typical age of NDD symptoms onset (Fig. 21).

Then, since it has been proposed that a more fruitful strategy to elucidate complex diseases such as DD/ID and ASD should include the examination of intermediate

phenotypes (Meyer-Lindenberg A and Weinberger DR, 2006), we deemed to perform a molecular screening of a 11,731 NDD patients' cohort by targeted resequencing our two best candidates: the *LZTR1* and the *ARVCF* genes. These genes were prioritized respectively due to the associated highest number of deleterious variants and, the highest CADD score retrieved in denovo-db, a previously described NDD database. Hence, since in this context intermediate phenotypes are meant to be biologically-based traits or mechanisms through which genes might affect behaviour, while "a genotype-first approach" constitutes a reverse strategy that assigns the pathogenic effects of many different genes and determines whether particular genotypes manifest as clinically recognizable phenotypes (Stessman HA et al., 2014; Jansen S et al., 2018), we aimed to demonstrate a high recurrence of severe *LZTR1* and *ARVCF* deleterious variants (i.e. LGD and/or MIS30) in our reference cohort.

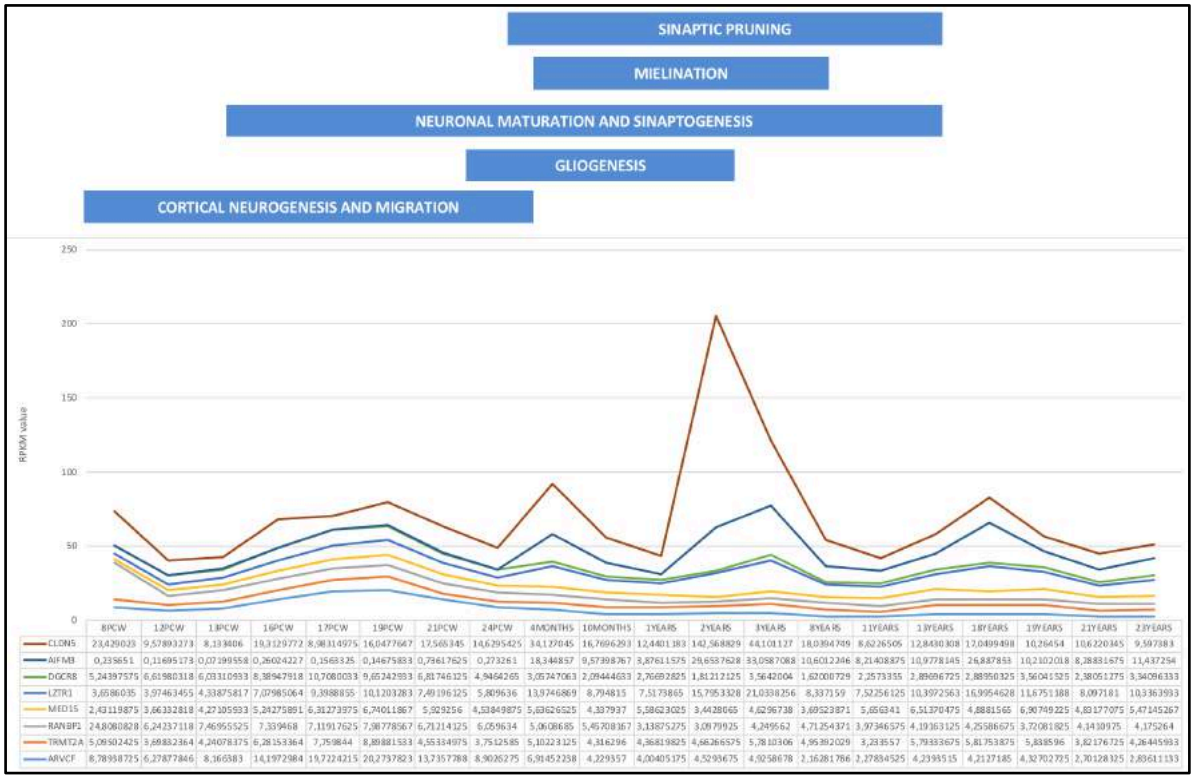


Fig. 21: Expression levels of the selected genes (coloured lines) in frontal cortex throughout development (from 8 post-conception weeks) up to 23 years. Above, blue rectangles depict neurodevelopment tome intervals. Under, in table, summary of data obtained from BrainSpan: Atlas of the Developing Human Brain. RPKM (Y, axis) values were max-min normalized across time points (X, axis).

As result, we validated 38 LGD/MIS30 QC-passing variants excluding those reported in ExAC and in dbSNP. Of these, we identified 26 novel LGD and MIS30 variants for the *LZTR1* and 12 for the *ARVCF* gene. Although this analysis has some limitations due to

the fact that it was possible to ascertain the rare private variant status only for some of them (see Tab. 5), the retrieved mutational burden in a large NDD cohort proved that these genes may: (i) contribute to the genetic background toward variability of neurodevelopmental phenotypes seen in individuals carrying 22q11.2 CNVs; (ii) constitute *per se* NDD driver genes.

In particular, ARVCF belongs to the p120ctn protein family which includes 4 products encoded by four independent genes, and distinct from the more ubiquitously expressed α- and β-catenins. Moreover, this protein family is constitutively expressed in central neurons, and it has recently become clear that exerts extensive functional roles in multiple aspects of neuronal morphology, synaptic structure, synaptic efficacy and molecular processes related to learning and memory (Seong E et al., 2015; Yuan L et al., 2017). Evolutionarily, these genes have been collectively generated from an ancient 'δ-catenin-like' gene and can be divided into two major classes, one of which comprises p0071 and δ-catenin, while the other includes p120ctn and ARVCF (Carnahan RH et al., 2010). The members of the p120ctn family share a common structural organization with some unique features; indeed, their structures show a series of ARM repeats included between N-terminal and C- terminal domains (Fig. 22).

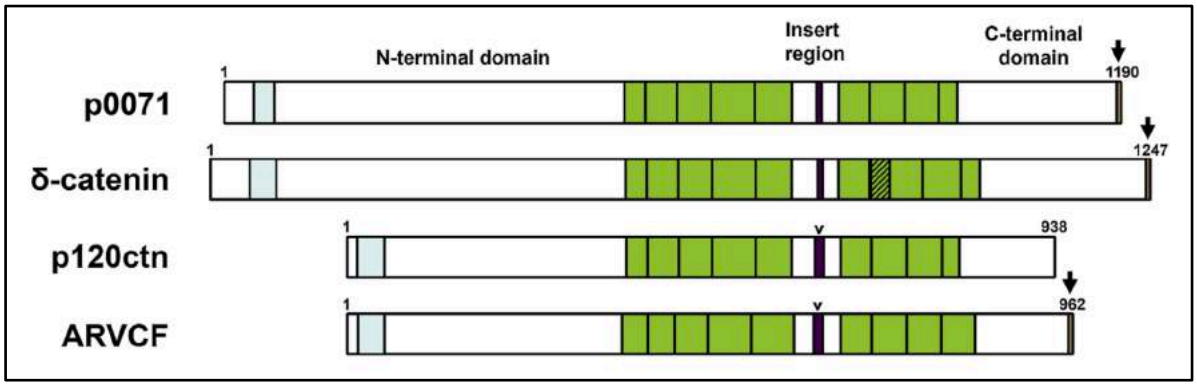


Figure 22: Ideogrammatic representation of the protein structures of the four members of the mouse p120ctn family, namely p0071, δ-catenin, p120ctn and ARVCF. The structures include an N-terminal domain, including a coiled-coiled domain (blue box), a series of ARM repeats (depicted in green boxes) and a C-terminal region. The polyK sequence is located within the insert region as indicated by black vertical bars. Number of amino acids for each protein is indicated by arrow above and at the end of each protein. The ARM repeats for each protein were defined by comparison to the ARM repeats of p120ctn after protein sequence alignment (adopted from Yuan L et al., 2017).

Moreover, while the C-terminal region of ARVCF, p0071 and δ-catenin also include a PDZ-binding motif that lacks in p120ctn, all these proteins carry a coiled-coiled domain at their N-terminal region suggesting that all members are unlikely to be completely

functionally redundant (Chauvet N et al. 2003; Markham NO et al., 2014; Yuan L et al., 2017). Finally a polyK sequence, resembling the nuclear localization signal and localized within the insert region between the ARM repeats, promotes their nuclear localization (Kelly KF et al., 2004; Ishiyama N et al., 2010). Indeed, the 962 amino acid ARVCF protein is also localized to the nucleus where it is supposed working as transcription regulator through its effectors (Sirotkin H et al., 1997; Bonné S et al. 1998; Mariner D et al., 2000; Rappe U et al., 2014). However, although insertions within the polyK sequence have been shown disrupting the intrinsic nuclear localization ability of p120ctn, these events do not impinge the ARVCF nuclear activity suggesting that also other determinants contribute to this family nuclear placement (Rappe U et al., 2014; Pieters, T et al. 2016; Yuan L et al., 2017). Of note, almost all the *ARCVF* deleterious variants found in our NDD cohort impinge the ARM protein domains that are supposed to be core to its molecular machinery since they are also variable conserved across multiple species (Fig. 23).

Figure 23: On the left, multiple sequence alignment of ARVCF protein orthologs around the sites of the mutations retrieved in our NDD cohort and whose events, highlighted and labelled within yellow vertical bars, mostly impinge invariant residues (*). On the right, ARVCF 3D structure depicting the c.1615C>T; p.(539*) mutation (green sphere) listed in denovo-db and leading to the highest CADD score of 42 described for this gene this in a patient affected by autism (SSC_14405.p1).

To date, several functional studies directly and/or indirectly suggest that *ARVCF* variations may resemble those neural impaired molecular pathways described for other p120ctn family members, as well as those pathogenetic mechanisms determined

by the dysfunction of similar adhesive-junction δ-catenin genes (Turner TN et al., 2015). Both Arvcf and p120ctn have been shown to extensively intervene in vertebrate embryogenesis sharing functional interplays with RhoA, Rac, and cadherin proteins in a neurodevelopmental context (Fang X et al., 2004). This may be due to the conservation of a RhoA-binding domain located into the insert region of these genes (Yuan L et al., 2017). Moreover, similar to p120ctn and δ-catenin, Arvcf influences dendrite-like process in NIH3T3 cells and, in *Xenopus* neural cells the same mechanism has been described triggered by an interaction with the KazrinA protein. This is another common effector that binds to ARVCF, δ-catenin and p0071 (Cho K. et al., 2010) and, since it constitutes a p190A RhoGAP binding partner that negatively regulates RhoA, the p120ctn family can be thought partly acting also through the Rho-signalling pathway. This was also confirmed in human cell lines where it has been shown that p120ctn can inhibit Rho activation due to the interaction between its C-terminal domain and the p190A RhoGAP (Zebda N et al., 2013). These evidences, taken together, indicate that the RhoGAP may be a common downstream effector for all the p120ctn family members (Yuan L et al., 2017).

However, while all the above-mentioned dendritic branching abilities deserve to be further and specifically tested for each gene in human brain, in reference to *ARVCF* it is noteworthy to report that its protein constitutes an FMRP target that mediates both *FMR1*-driven synaptic constitution and neurite outgrowth regulation (Nolze A et al., 2013; Yuan L et al., 2017). In this frame, and through high-throughput quantitative proteomic analysis, it was also been shown that Arvcf can be constantly found decreased in synaptic adherent junctions of *fmr1*-/- mouse cortical neurons impinging their neuron-neuron cadherin adhesion, which is a crucial determinant for synaptic establishment and maturation (Liao L et al., 2008). Thus, on the biological level, these functional interactions: (i) account for the deleterious events burden retrieved in our patients cohort, (ii) prove that this gene should be considered a novel NDD gene, and (iii) solve the previous *ARVCF* associations with the human psychiatric nosology since some of its haplotypes and/or other population genetics studies had over the time pointed out an increased schizophrenic relative risk (Sanders AR et al., 2005; Mas S et al., 2009; Mas S et al., 2010).

Conversely, although the *LZTR1* gene is already ranked as a confident NDD gene in SFARI due to the LGD heterozygotic private variants over-transmission seen in ASD probands families (Krumm N et al., 2015), its enrichment constituted an unexpected result both in our prioritisation analysis, and in the subsequent targeted resequencing screening due to the consistent mutational burden retrieved in our patients' cohort. Initially described as a putative transcriptional regulator based on weak homology to members of the basic leucine zipper-like family, the protein encoded by the *LZTR1* gene has subsequently been shown to localize exclusively to the Golgi network where it may help in stabilizing its complex (RefSeq, Jul 2008). To date, *LZTR1* germline mutations have been linked to the Noonan Syndrome, 10 disease (OMIM: 616564) accounting for both the autosomal, dominant or recessive, patterns of inheritance (Yamamoto GL et al., 2015; Ghedira N et al., 2017; Johnston JJ et al., 2018); moreover, its germline and somatic loss of function variants have been also respectively described predisposing to an inherited multiple schwannomatosis disorder (OMIM: 615670; Piotrowski A et al., 2014) or to other brain cancer diseases, i.e. the multiforme glioblastomatosis (Frattini V et al., 2013). The existence of either recessive or dominant forms of Noonan Syndrome (NS) associated with *LZTR1* variants has several implications for this gene pathogenic model (Johnston JJ et al., 2018). First, to date all the variants associated with the NS dominant form are missense variants (Suppl. Fig. 1 and Suppl. Tab. 16; Chen PC et al., 2014; Yamamoto GL et al., 2015; Ghedira N et al. 2017), whereas the NS recessive form was retrieved caused by either biallelic hypomorphic variants or compound heterozygosity for one hypomorphic mutation and one loss-of-function deleterious variants (Suppl. Fig. 2 and Suppl. Tab. 17; Biesecker LG, 2018). Interestingly, the mutational position might determine where LZTR1 variants can cause a dominant or a recessive NS form. Indeed, although the numbers of missense variants in dominant *LZTR1*-associated Noonan syndrome are small, they all reside between codons 119 and 287 within the series of Kelch domains, whereas the variants described in the recessive form span throughout the protein (Johnston JJ et al., 2018). Moreover, the *LZTR1* locus is harboured within the typical deleted region seen in the 22q11.2DS, and to our best knowledge, there are no NS patients carrying both the common 22q11.2 deletion in *trans* with an *LZTR1* SNV in the other allele. This also suggests that this locus haploinsufficiency could be tolerated and/or that it could at least contribute to the heterogeneity of phenotypes seen in

22q11.2DS. Hence, taking together these evidences, we can formulate the following two main hypotheses: (i) the reduction of LZTR1 function activity below 50% of its wild-type activity may result in NS either through dominant negative LZTR1 monoallelic pattern or the combination of hypomorphic and loss-of-function variants in the recessive form, (ii) *LZTR1* biallelic loss-of-function variants might be nonviable (Biesecker LG, 2018; Johnston JJ et al., 2018). However, even though these hypotheses could certainly contribute to improve our understanding on *LZTR1*-related pleiotropy, the LZTR1 pathogenicity model must also take into account its driven schwannomatosis susceptibility inasmuch as *LZTR1* germline mutations can be involved in a two- or three-step mutational process leading to tumor development through additional somatic hits that so far have been showed occurring on *NF2* and/or *SMARCB1* genes (Kehrer-Sawatzki H et al., 2017). Thus, to explain and to integrate in this model also the *LZTR1* NDD role highlighted by our analysis, we came up with the idea to widen the scope on the functional role played by its domains, Kelch and BTB. Indeed, clustering into a candidate gene the recurrence of LGD and MIS30 mutations has been recently proved to be a valid approach to highlight important, and even novel, functional NDD domains providing new insights into these disorders pathogenesis (Geisheker MR et al., 2017).

In this view, the human BTB-kelch protein leucine zipper-like transcriptional regulator 1 (LZTR1, OMIM: 600574) consists of 840 amino acids and contains six Kelch motifs and two BTB/POZ domains; however, in contrast to all other known BTB-Kelch-motif containing proteins, LZTR1 presents its Kelch motifs in the N terminus and its BTB/POZ domains in the C terminus, lacking of the BACK domains which appear in other BTB-kelch protein family members and constituting an unusual BTB-kelch protein (Kurahashi H et al., 1995; Nacak TG et al., 2006). The Kelch protein superfamily shows high structural as well as functional diversity, and it has been retrieved mostly interacting with the actin at the cytoplasmic level playing a role in cellular architecture and organization (Adams J et al., 2000). Instead, LZTR1 does not co-localize with actin, but it has been shown to localize to the Golgi complex marking a clear difference from all other BTB-Kelch proteins that exhibit a cytoplasmic distribution (Adams J et al., 2000; Nacak TG et al., 2006). As a consequence, LZTR1 could stabilize the Golgi complex as other BTB-Kelch proteins contribute to stabilize the cellular architecture

through the interaction with other structural proteins (Adams J et al., 2000; Nacak TG et al., 2006). Moreover, it has been also shown that the second LZTR1 BTB/POZ domain mediates the localization to the Golgi complex's cytoplasmic surface since the truncated mutant LZTR1-ΔBTB/POZ-2 loses this property (Nacak TG et al., 2006). So far, the BTB/POZ domain has been described mediating the protein's homomeric dimerization, and defining a recognition motif for the assembly of substrate-specific RING/cullin/BTB ubiquitin ligase complex (Geyer R et al. 2003; Kelly KF et al., 2004; Frattini V et al., 2013). Following these evidences, we explored the LZTR1 interactome network by using INstruct, a tool that combines the scale of available high-quality binary protein interaction data with the specificity of atomic-resolution structural information derived from co-crystal studies (Meyer MJ et al., 2013). The experiment result suggests that LZTR1: (i) may act as a homomeric dimer (Fig. 24a), (ii) may interact with other, yet less characterized protein pathways, trough its BTB domains i.e. with the ZBTB1 protein (Fig. 24b) which is a transcriptional repressor that belongs to the zinc finger BTB/POZ family (Liu Q et al., 2011; Kim H et al., 2014).

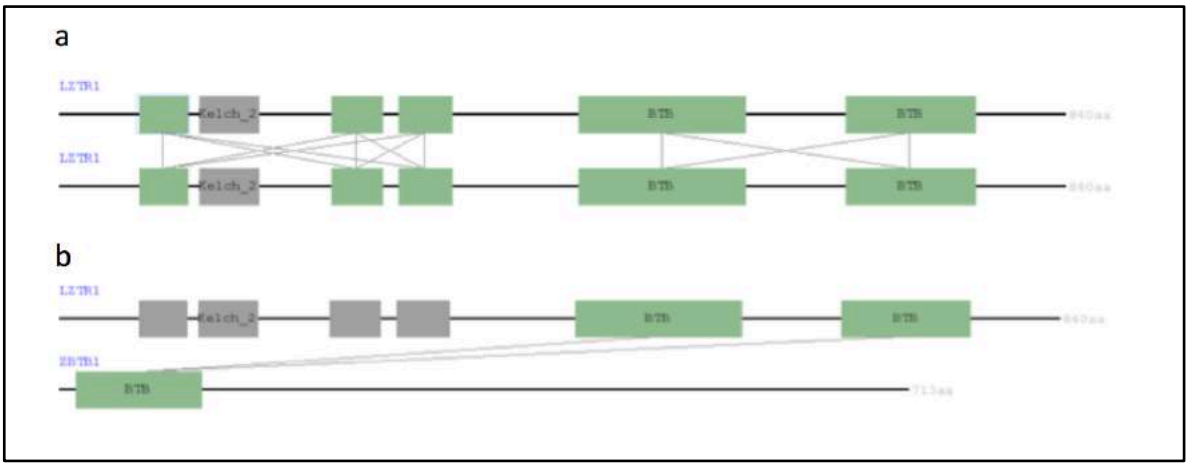


Figure 24: INstruct LZTR1 interactome network analysis result. LZTR1 may act as a homomeric dimer through its Kelch and BTB domains, and may interact with the ZBTB1 protein through their BTB domains. In figure: grey domains do not facilitate the given protein-protein interaction, whereas Interacting domains are shown in green; grey lines indicate a domain-domain interaction that is inferred using the INstruct homology-based inference method.

Although, on the molecular level, further studies should be undertaken to reveal the exact LZTR1 function, this first evidence of a homodimeric protein complex can more reasonably justify the existence of different NS inheritance patterns that, as predicted by other functional studies, do not fall into the canonical NS-associated RAS/MAPK signaling pathway (Ghedira N et al., 2017; Bustelo R et al., 2018).

In this view, while heterozygous -antimorphic- mutations may overrule the LZTR1 complex function activity arising whereby they can antagonistically affect the wild-type allele, the LZTR1 homodimer activity lowered at least under the 50% of its functionality threshold may finally explain the NS recessive pattern due to the occurrence of a different mutation in each locus according to the previously described combinations.

However in this model, where the reduction of the homodimer functional activity is directly proportional to the appearance of the NS typical features (Fig. 25), it should be taken into account also a yet unexplored BTB-domain driven role for both the NDDs and the ectodermic neoplasm predisposition. Indeed, not only several genes that contain the Broad complex, Tramtrack, and Bric-à-brac (BTB) transcriptional factor domains have been shown to play a critical role for learning, memory and behavior (Li W et al., 2004; Spletter ML et al., 2007; Jones KA et al. 2018). Specifically, the LZTR1 BTB domains have also been highlighted as adaptors of CUL3-containing E3 ligase complexes, whose activity is essential for the cortical neurogenesis (Dubiel W et al., 2017). Moreover, mutations targeting the BTB-domains on LZRT1-CUL3 binding sites determine stem cell features retained in neuroectodermal neoplasm lines maintaining their undifferentiated state (Frattini V et al., 2013). These evidences, taken together, support once again our hypothesis given also the fact that over two thirds of the mutations in our NDD patients were retrieved within, and/or included between, the LZTR1 BTB-domains (Fig. 26).

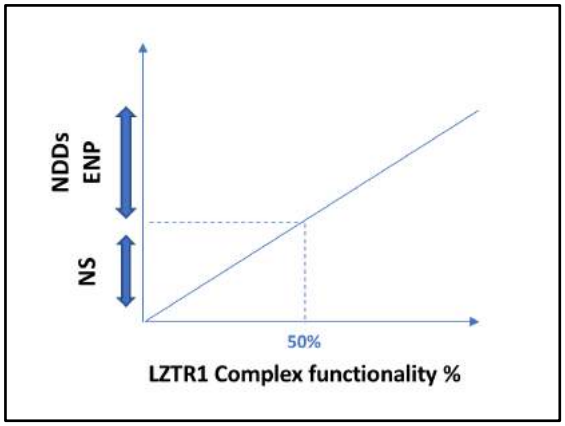


Figure 25: LZTR1 homodimeric protein complex's pathogenetic model predicting the NS features' appearance (Y, axis) below the theoretical 50% threshold (dashed blue line) of its overall activity (X, axis). This model allows also to speculate that events leading to variable complex impairment above the identified critical threshold may be responsible for neurodevelopmental disorders (NDDs) and/or may predispose to the ectodermic neoplasm (ENP).

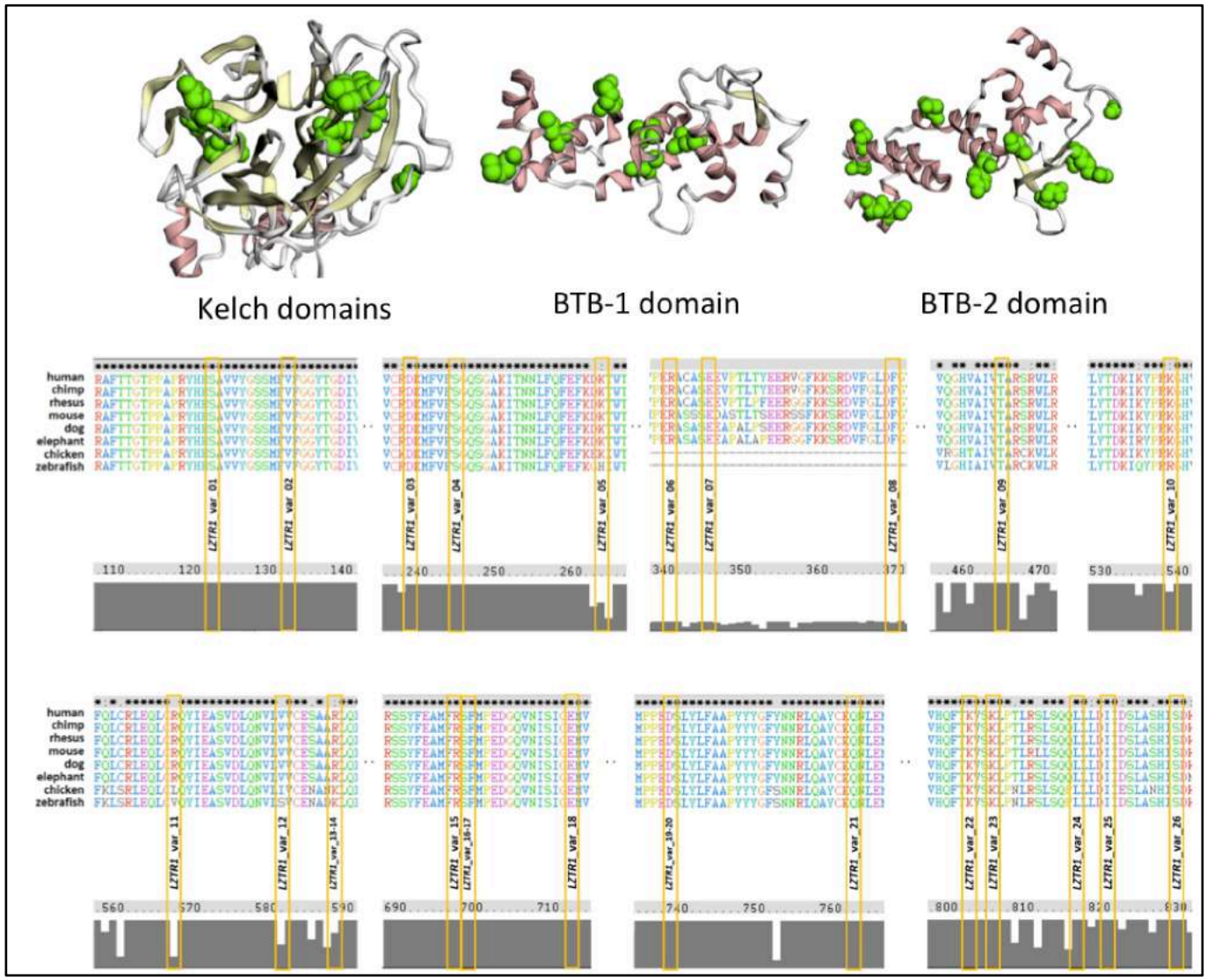


Figure 26: On the top, *LZTR1* LGD and MIS30 variants retrieved in our cohort (depicted by green spheres) mapped into LZTR1 protein domains according to their genomic position using MuPIT (http://mupit.icm.jhu.edu/MuPIT_Interactive) tool. Below, multiple sequence alignment of LZTR1 protein orthologs around the sites of the mutations retrieved in our NDD cohort and whose events, highlighted and labelled within yellow vertical bars, mostly impinge invariant residues (*). Of note, two thirds of these deleterious events map within, and/or are included between, the LZTR1 BTB-domains.

Finally, since no genes harboured within the distal 22q11.2 region reached nominal significance to all the enrichment criteria chosen for our genotype-first pass analysis, we developed a "phenotype-first approach" that could: (i) allow us to interface, anchor, and compare the phenotype features seen in this cohort to the genome, and (ii) contribute to identify novel NDD dosage-imbalance sensitive genes, whose pathogenetic molecular mechanisms were partially considered in the previous model's statistical assay. To this end, we built the only available distal 22q11.2 CNVs "co-morbidity" map collecting the HPO clinical information seen in distal 22q11.2 rearranged patients, describing a common phenotype broken out *per* single LCRs22 and *per* both deletion and duplication events (see Chapter 3.2).

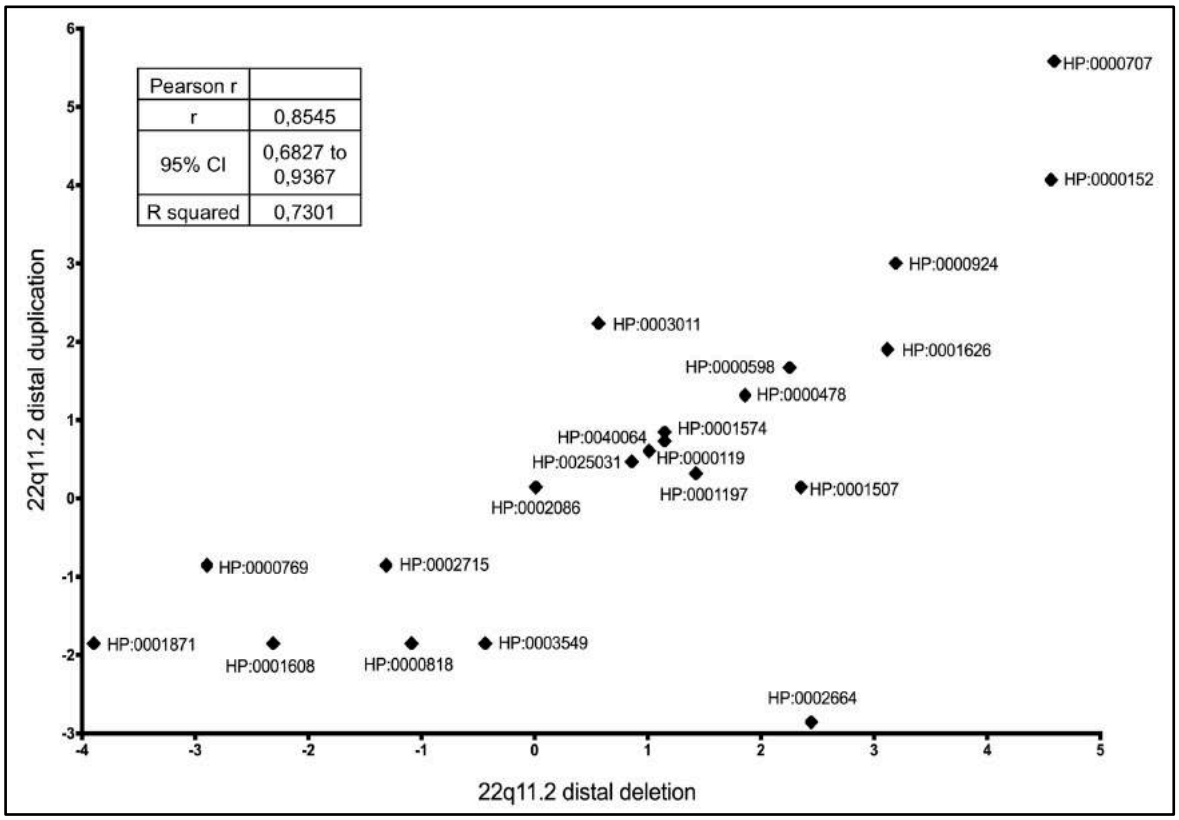


Figure 27: Log10 plot of nodes "0" HPO frequency % correlation by rearrangement type. Person r's test result describes that the congenital anomalies retrieved in the 22q11.2 cohort of rearranged patients mostly cluster into deletion events. HPO legend as previously reported in Fig. 12.

Following this workflow, we retrieved that the highest features' signatures are respectively clustered at the distal 22q11.2 D interval for deletion (55% of all the features retrieved, which represent an enrichment factor of 2.21 times higher than expected by a random chance assignment), and between the E to F intervals for duplication events (62.5% of all the features retrieved, enrichment factor of 1.25 times

higher than expected). Moreover, further analysing the nodes "0" HPO % frequency correlation by rearrangement type, we were able to describe that 22q11.2 distal deletion and duplication events respectively define a syndromic versus a non-syndromic neurodevelopmental disorder (Pearson r: 0,854; Fig. 27), especially due different sets of congenital cardiopathies, craniofacial dysmorphisms, and other congenital anomalies mostly clustered into deletion events, as opposed to ASD and/or DD/ID that constitute common shared features (Suppl. Fig. 3 and Suppl. Fig. 4).

Then, when these results were integrated with the previously described CNV morbidity map generated from the comparison of 29,085 NDD cases and 19,584 controls (see Chapter 3.3), we identified two small genomic frames of interest with both statistical significance and clinical relevance. As a matter of fact, we highlighted:
1) in the D interval, a ~300 Kb genomic frame ranging from 20,251,957 to 20,551,970 (NCBI36/hg18 coordinates) at the 22q11.21 sub-band which harbours the *UBE2L3, YDJC, CCDC116, SDF2L1, PPIL2, YPEL1* and *MAPK1* genes and, whose loci rearrangements are seen in 51 patients versus 10 in controls (One-tailed Fischer p-value: 5.418e-05, Odd Ratio: 3.438);
2) in the E interval, a ~256 Kb genomic frame ranging from 21,731,593 to 21,990,224 (NCBI36/hg18 coordinates) at the 22q11.23 sub-band which harbours the *RTDR1, GNAZ, RAB36* and *BCR* genes and, whose loci rearrangements are seen in 51 patients versus 8 controls (One-tailed Fischer p-value: 7.648e-06, Odd Ratio: 4.298).

Although we did not retrieve overlapping deletion events in the distal D genomic interval in a new enrolled NDD patients' cohort, we identified two proband carrying the same clinical features and an inherited 22q11.22q11.23 microduplication fully incorporating our predicted "telomeric" genomic frame of interest. This evidence, further suggesting a pivotal role of this region imbalance for the observed phenotype, prompted us to functionally study this duplication to uncover any causative pathogenic molecular mechanisms. The patients' SRO harbours 1 microRNA (*hsa-miR-650*) and 5 genes: *IGLL5, RTDR1* (gene alias *RSPH14*), *GNAZ, RAB36* and *BCR*. Since this SRO copy number gain is classified as an uncertain significance variation, our patients underwent an extensive diagnostic workup in accordance to the current guidelines that did not led us to recognize any known etiologic causes of intellectual disability

and/or seizures (van Karnebeek CD et al. 2014; Moeschler JB et al. 2014). Thus, due to the absence of gross functional imbalance, it appeared reasonable to assume that the neurological impairment observed in these patients was caused by an altered dosage effect of one or more of those genes encoded within. However, few information on the role of the aforementioned genes in humans are available. The only coding genes extensively so far studied is *BCR* since its locus is involved in the reciprocal translocation between chromosomes 22 and 9, which produces the Philadelphia chromosome often found in patients with chronic myelogenous leukaemia (Rafiei A et al., 2015). *hsa-miR-650* up-regulation seems to correlate with hepatocellular carcinoma pathogenesis and its disruptions have been implicated in gastric and colorectal cancer tumorigenicity (Zeng Z et al., 2013). The *IGLL5* gene, located within the immunoglobulin lambda locus, encodes one of the immunoglobulin lambda-like polypeptides (Guglielmi P and Davi F, 1991). The *RTDR1* gene (alias *RSPH14*) was described deleted in paediatric rhabdoid tumors and its protein is known to have a slight similarity to yeast vacuolar proteins (Zhou JY et al., 2000). The protein encoded by the *GNAZ* gene is a member of a G protein subfamily that mediates signal transduction in pertussis toxin-insensitive systems that could play a role in the balance of perilymphatic and endolymphatic cochlear fluids (Magovcevic I et al., 1995). Finally, the RAB36 protein is a less characterized RAB-family protein mostly conserved in vertebrates an involved in neuronal development through the RAB35 pathway (Kobayashi H et al., 2014).

Thus, in order to highlight the best candidate, as a first step the expression of genes encoded within the SRO was verified through a Real Time assay confirming all their constitutive brain expression, even though with different levels. Hence, comparing the 5 brain expressed genes (*RTDR1, GNAZ, BCR, RAB36* and *IGGL5*) between the two patients, their carrying parents and a negative controls' set, we disclosed a specific and significant altered RAB36 imbalance suggesting an altered gene-dosage effect for this gene that emerged as the best candidate effector. In this view, RAB GTPases are emerging for their crucial role in compartment specific directional control of vesicles formation, transport and fusion, playing a prominent role in brain development and maturation (Mignogna ML et al., 2015). Thus, as the function of endogenous RAB36 protein is thought to be involved in vesicular trafficking, its imbalance may interfere in

neurotransmitter processing, causing an impaired trafficking mechanism of release or other forms of neural secretion and maturation. Although further functional studies are certainly necessary to better characterize the prosed pathogenic mechanism at the molecular level, our functional analysis through the GFP-Rab36 overexpression in primary E18 mouse hippocampal neuronal cultures showed a morphological maturation impairment due to lower dendrite branching (see Chapter 3.4 results) confirming to date that RAB36 may play a key role in NDD pathogenesis acting in a dosage-dependent manner.

CONCLUSION

In conclusion, in this research project we applied a genotype- and a phenotype-based approaches to further define, hence to experimentally confirm, the pivotal neurodevelopmental role of the 22q11.2 region. Through our genotype-first approach we identified 8 genes (*ARVCF*, *CLDN5*, *DGCR8*, *RANBP1*, *TRMT2A*, *MED15*, *AIFM3* and, *LZTR1*) that, according to our enrichment analysis criteria, can exert a downstream functional convergence upon those circuits involved in physiological neurodevelopment as well as in pathophysiological behaviors. Among them, the mutational burden retrieved in a large NDD patients' cohort by targeted resequencing *ARVCF* and *LZTR1*, selected as the two prioritized genes, proved that: (i) their impairment can contribute to the genetic background toward variability of neurodevelopmental phenotypes seen in individuals carrying 22q11.2 CNVs; (ii) these genes may constitute *per se* NDD driver.

Hence, in order to identify novel NDD genes potentially sensitive to dosage imbalance within the distal 22q11.2 chromosomal region, we developed a phenotype-first approach through the building of the distal 22q11.2 CNVs "co-morbidity" map that allowed us: (i) to elucidate that the 22q11.2 distal deletion and duplication events respectively define a syndromic versus a non-syndromic neurodevelopmental disorder; (ii) to identify two small critical genomic frames of interest within this region. Deep investigating two 22q11.22q11.23 carriers' families, whose SRO was retrieved fully incorporating one of our predicted pathogenic frames, we functionally showed that the *RAB36* gene may play a key role in NDD pathogenesis acting in a dosage-dependent manner.

Overall, this research project's results added new insights on the elucidation of 22q11.2DS/22q11.2DupS neurobehavioral outcomes as well as of NDDs' neuroscience physiopathology, widening the scope on those dynamics that constantly shape and re-shape the human brain.

ACKNOWLEDGMENT

I have many people to whom I would like to express my acknowledgement. First and foremost, I would like to thank Evan Eichler PhD and the postdoctoral researchers Arvis Sulovaris PhD and Tianyun Wang PhD at Eichler Lab, Department of Genome Sciences, University of Washington, USA for having put my training and education in touch with the forefront of human genetics research. Their schooling and friendship will have a lasting positive impact on me for many more years to come.

I am very grateful to my PhD advisors committee Antonio Pizzuti MD, PhD and Viviana Caputo PhD for ensuring that I became the best physician that I could during the program of the Medical Genetics PhD School at Sapienza University of Rome, Italy. I would like to thank also my colleagues Agnese Giovannetti BS, Alice Traversa PhD and Maria Luce Genovesi BS for their continuous help and support during this journey. I am deeply beholden to Patrizia D'Adamo PhD and Maria Lidia Mignogna PhD of the Molecular Genetics of Intellectual Disabilities Unit, San Raffaele Scientific Institute of Milan, Italy and to Aldo Nicosia PhD at the IAMC-CNR Palermo, Italy for their collaboration in carrying out part of the functional studies presented in this study. I would like to thank Maria Piccione MD of the Regional Referral Centre for Rare Genetic and Chromosomal Diseases, AOOR "Villa Sofia - Cervello" Hospital, Palermo, Italy that, before and above anything else, taught me the true passion for this discipline.

Finally, I would like to dedicate this work to the children who took part in its research, to all those children that I met over the time because of my profession, to their families, to my family and especially to Emanuela. I am truly indebted to all of them for continuously teaching me to look thoroughly into everything, near and afar, alike in matter and in cause, towards the real meanings of life and love.

BIBLIOGRAPHY

Adams, J., Kelso, R., & Cooley, L. (2000). The kelch repeat superfamily of proteins: propellers of cell function. Trends in cell biology, 10(1), 17-24.

Aizawa, H. (2013). Habenula and the asymmetric development of the vertebrate brain. Anatomical science international, 88(1), 1-9.

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., ... & Durkin, M. S. (2018). Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveillance Summaries, 67(6), 1.

Bassett, A. S., Caluseriu, O., Weksberg, R., Young, D. A., & Chow, E. W. (2007). Catechol-O-methyl transferase and expression of schizophrenia in 73 adults with 22q11 deletion syndrome. Biological Psychiatry, 61(10), 1135-1140.

Bassett, A. S., Lowther, C., Merico, D., Costain, G., Chow, E. W., Van Amelsvoort, T., ... & Murphy, K. (2017). Rare genome-wide copy number variation and expression of schizophrenia in 22q11. 2 deletion syndrome. American Journal of Psychiatry, 174(11), 1054-1063.

Beddow, R. A., Smith, M., Kidd, A., Corbett, R., & Hunter, A. G. (2011). Diagnosis of distal 22q11. 2 deletion syndrome in a patient with a teratoid/rhabdoid tumour. European journal of medical genetics, 54(3), 295-298.

Ben-Shachar, S., Ou, Z., Shaw, C. A., Belmont, J. W., Patel, M. S., Hummel, M., ... & Lalani, S. R. (2008). 22q11. 2 distal deletion: a recurrent genomic disorder distinct from DiGeorge syndrome and velocardiofacial syndrome. The American Journal of Human Genetics, 82(1), 214-221.

Bender, H. U., Almashanu, S., Steel, G., Hu, C. A., Lin, W. W., Willis, A., ... & Valle, D. (2005). Functional consequences of PRODH missense mutations. The American Journal of Human Genetics, 76(3), 409-420.

Bertocci, B., Miggiano, V., Da Prada, M., Dembic, Z., Lahm, H. W., & Malherbe, P. (1991). Human catechol-O-methyltransferase: cloning and expression of the membrane-associated form. Proceedings of the National Academy of Sciences, 88(4), 1416-1420.

Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A., & Cairns, M. J. (2010). Schizophrenia is associated with an increase in cortical microRNA biogenesis. Molecular psychiatry, 15(12), 1176.

Biesecker, L. G. (2018). Response to Nakaguma et al. Genetics in Medicine, 1.

Bonné, S., van Hengel, J., & Van Roy, F. (1998). Chromosomal Mapping of Human Armadillo Genes Belonging to the p120 ctn/Plakophilin Subfamily. Genomics, 51(3), 452-454.

Boot, E., & van Amelsvoort, T. A. M. J. (2012). Neuroimaging correlates of 22q11. 2 deletion syndrome: implications for schizophrenia research. Current topics in medicinal chemistry, 12(21), 2303-2313.

Bourdeaut, F., Lequin, D., Brugières, L., Reynaud, S., Dufour, C., Doz, F., ... & Orbach, D. (2011). Frequent hSNF5/INI1 germline mutations in patients with rhabdoid tumor. Clinical Cancer Research, 17(1), 31-38.

Bosse, K. R., Shukla, A. R., Pawel, B., Chikwava, K. R., Santi, M., Tooke, L., ... & Bagatell, R. (2014). Malignant rhabdoid tumor of the bladder and ganglioglioma in a 14 year-old male with a germline 22q11. 2 deletion. Cancer genetics, 207(9), 415-419.

Bucan, M., Abrahams, B. S., Wang, K., Glessner, J. T., Herman, E. I., Sonnenblick, L. I., ... & Kim, C. (2009). Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes. PLoS genetics, 5(6), e1000536.

Burnside, R. D. (2015). 22q11. 21 deletion syndromes: a review of proximal, central, and distal deletions and their associated features. Cytogenetic and genome research, 146(2), 89-99.

Bustelo, X. R., Crespo, P., Fernández-Pisonero, I., & Rodríguez-Fdez, S. (2018). RAS GTPase-dependent pathways in developmental diseases: old guys, new lads, and current challenges. Current opinion in cell biology, 55, 42-51.

Bruce, S., Hannula-Jouppi, K., Puoskari, M., Fransson, I., Simola, K. O., Lipsanen-Nyman, M., & Kere, J. (2009). Submicroscopic genomic alterations in Silver-Russell syndrome and Silver-Russell-like patients. Journal of medical genetics.

Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., ... & Guan, P. (2015). A novel approach to high-quality postmortem tissue procurement: the GTEx project. Biopreservation and biobanking, 13(5), 311-319.

Carvalho, M. R. S., Vianna, G., Oliveira, L. D. F. S., Costa, A. J., Pinheiro‐Chagas, P., Sturzenecker, R., ... & Haase, V. G. (2014). Are 22q11. 2 distal deletions associated with math difficulties?. American Journal of Medical Genetics Part A, 164(9), 2256-2262.

Carnahan, R. H., Rokas, A., Gaucher, E. A., & Reynolds, A. B. (2010). The molecular evolution of the p120-catenin subfamily and its functional associations. PLoS One, 5(12), e15747.

Chakrapani, A. L., White, C. R., Korcheva, V., White, K., Lofgren, S., Zonana, J., ... & Mansoor, A. (2012). Congenital extrarenal malignant rhabdoid tumor in an infant with distal 22q11. 2 deletion syndrome: the importance of SMARCB1. The American Journal of Dermatopathology, 34(6), e77-e80.

Chang, J., Zhao, L., Chen, C., Peng, Y., Xia, Y., Tang, G., ... & Mei, L. (2015). Pachygyria, seizures, hypotonia, and growth retardation in a patient with an atypical 1.33 Mb inherited microduplication at 22q11. 23. Gene, 569(1), 46-50.

Chauvet, N., Prieto, M., Fabre, C., Noren, N. K., & Privat, A. (2003). Distribution of p120 catenin during rat brain development:: potential role in regulation of cadherin-mediated adhesion and actin cytoskeleton organization. Molecular and Cellular Neuroscience, 22(4), 467-486.

Chen, P. C., Yin, J., Yu, H. W., Yuan, T., Fernandez, M., Yung, C. K., ... & Morgan, M. B. (2014). Next-generation sequencing identifies rare variants associated with Noonan syndrome. Proceedings of the National Academy of Sciences, 111(31), 11473-11478.

Chen, J., Lipska, B. K., Halim, N., Ma, Q. D., Matsumoto, M., Melhem, S., ... & Egan, M. F. (2004). Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. The American Journal of Human Genetics, 75(5), 807-821.

Cho, K., Vaught, T. G., Ji, H., Gu, D., Papasakelariou-Yared, C., Horstmann, N., ... & Reynolds, A. B. (2010). Xenopus Kazrin interacts with ARVCF-catenin, spectrin and p190B RhoGAP, and modulates RhoA activity and epithelial integrity. Journal of cell science, jcs-072041.

Clements, C. C., Wenger, T. L., Zoltowski, A. R., Bertollo, J. R., Miller, J. S., de Marchena, A. B., ... & Emanuel, B. S. (2017). Critical region within 22q11. 2 linked to higher rate of autism spectrum disorder. Molecular autism, 8(1), 58.

Coe, B. P., Witherspoon, K., Rosenfeld, J. A., Van Bon, B. W., Vulto-van Silfhout, A. T., Bosco, P., ... & Schuurs-Hoeijmakers, J. H. (2014). Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nature genetics, 46(10), 1063.

Cooper, G. M., Coe, B. P., Girirajan, S., Rosenfeld, J. A., Vu, T. H., Baker, C., ... & Abdel-Hamid, H. (2011). A copy number variation morbidity map of developmental delay. Nature genetics, 43(9), 838.

Coppinger, J., McDonald-McGinn, D., Zackai, E., Shane, K., Atkin, J. F., Asamoah, A., ... & Feldman, H. (2009). Identification of familial and de novo microduplications of 22q11. 21–q11. 23 distal to the 22q11. 21 microdeletion syndrome region. Human molecular genetics, 18(8), 1377-1383.

Costain, G., Lionel, A. C., Merico, D., Forsythe, P., Russell, K., Lowther, C., ... & Chow, E. W. (2013). Pathogenic rare copy number variants in community-based schizophrenia suggest a potential role for clinical microarrays. Human molecular genetics, 22(22), 4485-4501.

Christian, S. L., Brune, C. W., Sudi, J., Kumar, R. A., Liu, S., Karamohamed, S., ... & Gergel, J. (2008). Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biological psychiatry, 63(12), 1111-1117.

D'Angelo, C. S., Varela, M. C., de Castro, C. I., Kim, C. A., Bertola, D. R., Lourenço, C. M., ... & Koiffmann, C. P. (2014). Investigation of selected genomic deletions and duplications in a cohort of 338 patients presenting with syndromic obesity by multiplex ligation-dependent probe amplification using synthetic probes. Molecular cytogenetics, 7(1), 75.

Demily, C., Lesca, G., Poisson, A., Till, M., Barcia, G., Chatron, N., ... & Munnich, A. (2018). Additive Effect of Variably Penetrant 22q11. 2 Duplication and Pathogenic Mutations in Autism Spectrum Disorder: To Which Extent Does the Tree Hide He Forest?. Journal of autism and developmental disorders, 1-4.

Descartes, M., Franklin, J., de Ståhl, T. D., Piotrowski, A., Bruder, C. E., Dumanski, J. P., ... & Mikhail, F. M. (2008). Distal 22q11. 2 microduplication encompassing the BCR gene. American Journal of Medical Genetics Part A, 146(23), 3075-3081.

Digilio, M. C., Versacci, P., Bernardini, L., Novelli, A., Marino, B., & Dallapiccola, B. (2015). Left ventricular non compaction with aortic valve anomalies: a recurrent feature of 22q11. 2 distal deletion syndrome. European journal of medical genetics, 58(8), 406-408.

Dougherty, J. D., Schmidt, E. F., Nakajima, M., & Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic acids research, 38(13), 4218-4230.

Driscoll, D. A., Salvin, J., Sellinger, B., Budarf, M. L., McDonald-McGinn, D. M., Zackai, E. H., & Emanuel, B. S. (1993). Prevalence of 22q11 microdeletions in DiGeorge and velocardiofacial syndromes: implications for genetic counselling and prenatal diagnosis. Journal of Medical Genetics, 30(10), 813-817.

Dubiel, W., Dubiel, D., Wolf, D. A., & Naumann, M. (2017). Cullin 3-Based Ubiquitin Ligases as Master Regulators of Mammalian Cell Differentiation. Trends in biochemical sciences.

Eaton, K. W., Tooke, L. S., Wainwright, L. M., Judkins, A. R., & Biegel, J. A. (2011). Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. Pediatric blood & cancer, 56(1), 7-15.

Fagerberg, C. R., Graakjaer, J., Heinl, U. D., Ousager, L. B., Dreyer, I., Kirchhoff, M., ... & Sorensen, K. (2013). Heart defects and other features of the 22q11 distal deletion syndrome. European journal of medical genetics, 56(2), 98-107.

Fang, X., Ji, H., Kim, S. W., Park, J. I., Vaught, T. G., Anastasiadis, P. Z., ... & McCrea, P. D. (2004). Vertebrate development requires ARVCF and p120 catenins and their interplay with RhoA and Rac. J Cell Biol, 165(1), 87-98.

Fournier, A. E., GrandPre, T., & Strittmatter, S. M. (2001). Identification of a receptor mediating Nogo-66 inhibition of axonal regeneration. Nature, 409(6818), 341.

Frattini, V., Trifonov, V., Chan, J. M., Castano, A., Lia, M., Abate, F., ... & Danussi, C. (2013). The integrated landscape of driver genomic alterations in glioblastoma. Nature genetics, 45(10), 1141.

Fung, W. L. A., Butcher, N. J., Costain, G., Andrade, D. M., Boot, E., Chow, E. W., ... & García-Miñaúr, S. (2015). Practical guidelines for managing adults with 22q11. 2 deletion syndrome. Genetics in Medicine, 17(8), 599.

Geisheker, M. R., Heymann, G., Wang, T., Coe, B. P., Turner, T. N., Stessman, H. A., ... & Liebelt, J. (2017). Hotspots of missense mutation identify neurodevelopmental disorder genes and functional domains. Nature neuroscience, 20(8), 1043.

Geyer, R., Wee, S., Anderson, S., Yates, J., & Wolf, D. A. (2003). BTB/POZ domain proteins are putative substrate adaptors for cullin 3 ubiquitin ligases. Molecular cell, 12(3), 783-790.

Ghedira, N., Kraoua, L., Lagarde, A., Abdelaziz, R. B., & Olschwang, S. (2017). Further Evidence for the Implication of LZTR1, a Gene Not Associated with the Ras-Mapk Pathway, in the Pathogenesis of Noonan Syndrome. Biol Med (Aligarh), 9(414), 2.

Giannandrea, M., Bianchi, V., Mignogna, M. L., Sirri, A., Carrabino, S., D'Elia, E., ... & Ropers, H. H. (2010). Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. The American Journal of Human Genetics, 86(2), 185-195.

Gothelf, D., Schneider, M., Green, T., Debbané, M., Frisch, A., Glaser, B., ... & Eliez, S. (2013). Risk factors and the evolution of psychosis in 22q11. 2 deletion syndrome: a longitudinal 2-site study. Journal of the American Academy of Child & Adolescent Psychiatry, 52(11), 1192-1203.

Gothelf, D., Eliez, S., Thompson, T., Hinard, C., Penniman, L., Feinstein, C., ... & Morris, M. A. (2005). COMT genotype predicts longitudinal cognitive decline and psychosis in 22q11. 2 deletion syndrome. Nature neuroscience, 8(11), 1500.

Greene, C., Kealy, J., Humphries, M. M., Gong, Y., Hou, J., Hudson, N., ... & Grant, G. A. (2017). Dose-dependent expression of claudin-5 is a modifying factor in schizophrenia. Molecular psychiatry.

Green, T., Gothelf, D., Glaser, B., Debbane, M., Frisch, A., Kotler, M., ... & Eliez, S. (2009). Psychiatric disorders and intellectual functioning throughout development in velocardiofacial (22q11. 2 deletion) syndrome. Journal of the American Academy of Child & Adolescent Psychiatry, 48(11), 1060-1068.

Grati, F. R., Molina Gomes, D., Ferreira, J. C. P. B., Dupont, C., Alesi, V., Gouas, L., ... & Piotrowski, K. (2015). Prevalence of recurrent pathogenic microdeletions and microduplications in over 9500 pregnancies. Prenatal diagnosis, 35(8), 801-809.

Guglielmi, P., & Davi, F. (1991). Expression of a novel type of immunoglobulin Cλ transcripts in human mature B lymphocytes producing χ light chains. European journal of immunology, 21(2), 501-508.

Gulsuner, S., & McClellan, J. M. (2014). De novo mutations in schizophrenia disrupt genes co-expressed in fetal prefrontal cortex. Neuropsychopharmacology, 39(1), 238.

Hantash, F. M., Wang, B. T., Owen, R., Ross, L. P., Mahon, L. W., Boyar, F. Z., ... & Strom, C. M. (2012). Inherited and de novo 22q11. 2 distal duplications in two patients with autistic features, speech delay and no dysmorphology. Journal of pediatric genetics, 1(2), 115-124.

Hidding, E., Swaab, H., de Sonneville, L. M. J., van Engeland, H., & Vorstman, J. A. S. (2016). The role of COMT and plasma proline in the variable penetrance of autistic spectrum symptoms in 22q11. 2 deletion syndrome. Clinical genetics, 90(5), 420-427.

Hiatt, J. B., Pritchard, C. C., Salipante, S. J., O'Roak, B. J., & Shendure, J. (2013). Single molecule molecular inversion probes for targeted, high accuracy detection of low frequency variation. Genome research, gr-147686.

Hoeffding, L. K., Trabjerg, B. B., Olsen, L., Mazin, W., Sparsø, T., Vangkilde, A., ... & Werge, T. (2017). Risk of psychiatric disorders among individuals with the 22q11. 2 deletion or duplication: a Danish nationwide, register-based study. JAMA psychiatry, 74(3), 282-290.

Ishiyama, N., Lee, S. H., Liu, S., Li, G. Y., Smith, M. J., Reichardt, L. F., & Ikura, M. (2010). Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. Cell, 141(1), 117-128.

Jackson, E. M., Shaikh, T. H., Gururangan, S., Jones, M. C., Malkin, D., Nikkel, S. M., ... & Biegel, J. A. (2007). High-density single nucleotide polymorphism array analysis in patients with germline deletions of 22q11. 2 and malignant rhabdoid tumor. Human genetics, 122(2), 117-127.

Jansen, S., Hoischen, A., Coe, B. P., Carvill, G. L., Esch, H., Bosch, D. G., ... & Bon, B. W. (2018). A genotype-first approach identifies an intellectual disability-overweight syndrome caused by PHIP haploinsufficiency. European Journal of Human Genetics, 26(1), 54.

Jaffe, A. E., Straub, R. E., Shin, J. H., Tao, R., Gao, Y., Collado-Torres, L., ... & Colantuoni, C. (2018). Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. Nat Neurosci, 21, 1117-25.

Johnston, J. J., van der Smagt, J. J., Rosenfeld, J. A., Pagnamenta, A. T., Alswaid, A., Baker, E. H., ... & Dung, V. C. (2018). Autosomal recessive Noonan syndrome associated with biallelic LZTR1 variants. Genetics in Medicine.

Jones, K. A., Luo, Y., Dukes-Rimsky, L., Srivastava, D. P., Koul-Tewari, R., Russell, T. A., ... & Penzes, P. (2018). Neurodevelopmental disorder-associated ZBTB20 gene variants affect dendritic and synaptic structure. PloS one, 13(10), e0203760.

Kaufman, C. S., Genovese, A., & Butler, M. G. (2016). Deletion of TOP3B is associated with cognitive impairment and facial dysmorphism. Cytogenetic and genome research, 150(2), 106-111.

Kaminsky, E.B. et al. An evidence-based approach to establish the functional and clinical significance of copy number variants in intellectual and developmental disabilities. Genet. Med. 13, 777–784 (2011).

Kelly, K. F., Otchere, A. A., Graham, M., & Daniel, J. M. (2004). Nuclear import of the BTB/POZ transcriptional regulator Kaiso. Journal of cell science, 117(25), 6143-6152.

Kelly, K. F., Spring, C. M., Otchere, A. A., & Daniel, J. M. (2004). NLS-dependent nuclear localization of p120ctn is necessary to relieve Kaiso-mediated transcriptional repression. Journal of cell science, 117(13), 2675-2686.

Kim, H., Dejsuphong, D., Adelmant, G., Ceccaldi, R., Yang, K., Marto, J. A., & D'Andrea, A. D. (2014). Transcriptional repressor ZBTB1 promotes chromatin remodeling and translesion DNA synthesis. Molecular cell, 54(1), 107-118.

Kircher, M., Witten, D. M., Jain, P., O'roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics, 46(3), 310.

Kehrer-Sawatzki, H., Farschtschi, S., Mautner, V. F., & Cooper, D. N. (2017). The molecular pathogenesis of schwannomatosis, a paradigm for the co-involvement of multiple tumour suppressor genes in tumorigenesis. Human genetics, 136(2), 129-148.

Kirov, G., Rees, E., Walters, J. T., Escott-Price, V., Georgieva, L., Richards, A. L., ... & McCarroll, S. A. (2014). The penetrance of copy number variations for schizophrenia and developmental delay. Biological psychiatry, 75(5), 378-385.

Kobayashi H, Etoh K, Ohbayashi N, Fukuda M. 2014. Rab35 promotes the recruitment of Rab8, Rab13 and Rab36 to recycling endosomes through MICAL-L1 during neurite outgrowth. Biol Open, 3(9), 803-814.

Köhler, S., Vasilevsky, N. A., Engelstad, M., Foster, E., McMurry, J., Aymé, S., ... & Brudno, M. (2016). The human phenotype ontology in 2017. Nucleic acids research, 45(D1), D865-D876.

Kragness, S., Harrison, M. A., Westmoreland, J. J., Burstain, A., & Earls, L. R. (2018). Age-dependent expression pattern in the mammalian brain of a novel, small peptide encoded in the 22q11. 2 deletion syndrome region. Gene Expression Patterns, 28, 95-103.

Kurahashi, H., Akagi, K., Inazawa, J., Ohta, T., Niikawa, N., Kayatani, F., ... & Nishisho, I. (1995). Isolation and characterization of a novel gene deleted in DiGeorge syndrome. Human molecular genetics, 4(4), 541-549.

Lafay‐Cousin, L., Payne, E., Strother, D., Chernos, J., Chan, M., & Bernier, F. P. (2009). Goldenhar phenotype in a child with distal 22q11. 2 deletion and intracranial atypical

teratoid rhabdoid tumor. American Journal of Medical Genetics Part A, 149(12), 2855-2859.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., ... & Tukiainen, T. (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature, 536(7616), 285.

Li, W., Wang, F., Menut, L., & Gao, F. B. (2004). BTB/POZ-zinc finger protein abrupt suppresses dendritic branching in a neuronal subtype-specific and dosage-dependent manner. Neuron, 43(6), 823-834.

Liao, L., Park, S. K., Xu, T., Vanderklish, P., & Yates, J. R. (2008). Quantitative proteomic analysis of primary neurons reveals diverse changes in synaptic protein content in fmr1 knockout mice. Proceedings of the National Academy of Sciences, 105(40), 15281-15286.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., ... & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS medicine, 6(7), e1000100.

Lin, A., Ching, C. R., Vajdi, A., Sun, D., Jonas, R. K., Jalbrzikowski, M., ... & Dokoru, D. (2017). Mapping 22q11. 2 gene dosage effects on brain morphometry. Journal of Neuroscience, 3759-16.

Lopez-Rivera, E., Liu, Y. P., Verbitsky, M., Anderson, B. R., Capone, V. P., Otto, E. A., ... & Fasel, D. A. (2017). Genetic drivers of kidney defects in the DiGeorge syndrome. New England Journal of Medicine, 376(8), 742-754.

Lindgren, V., McRae, A., Dineen, R., Saulsberry, A., Hoganson, G., & Schrift, M. (2015). Behavioral abnormalities are common and severe in patients with distal 22q11. 2 microdeletions and microduplications. Molecular genetics & genomic medicine, 3(4), 346-353.

Liu, Q., Yao, F., Wang, M., Zhou, B., Cheng, H., Wang, W., ... & Wang, J. C. (2011). Novel human BTB/POZ domain-containing zinc finger protein ZBTB1 inhibits transcriptional activities of CRE. Molecular and cellular biochemistry, 357(1-2), 405-414.

Madan, S., Madan‐Khetarpal, S., Park, S. C., Surti, U., Bailey, A. L., McConnell, J., & Tadros, S. S. (2010). Left ventricular non‐compaction on MRI in a patient with 22q11. 2 distal deletion. American Journal of Medical Genetics Part A, 152(5), 1295-1299.

Magovcevic, I., Khetarpal, U., Bieber, F. R., & Morton, C. C. (1995). GNAZ in human fetal cochlea: expression, localization, and potential role in inner ear function. Hearing research, 90(1-2), 55-64.

Malhotra, D., & Sebat, J. (2012). CNVs: harbingers of a rare variant revolution in psychiatric genetics. Cell, 148(6), 1223-1241.

Maynard, T. M., Haskell, G. T., Peters, A. Z., Sikich, L., Lieberman, J. A., & LaMantia, A. S. (2003). A comprehensive analysis of 22q11 gene expression in the developing and adult brain. Proceedings of the National Academy of Sciences, 100(24), 14433-14438.

Mariner, D. J., Wang, J., & Reynolds, A. B. (2000). ARVCF localizes to the nucleus and adherens junction and is mutually exclusive with p120 (ctn) in E-cadherin complexes. J Cell Sci, 113(8), 1481-1490.

Markham, N. O., Doll, C. A., Dohn, M. R., Miller, R. K., Yu, H., Coffey, R. J., ... & Reynolds, A. B. (2014). DIPA-family coiled-coils bind conserved isoform-specific head domain of p120-catenin family: potential roles in hydrocephalus and heterotopia. Molecular biology of the cell, 25(17), 2592-2603.

Marshall, C. R., Howrigan, D. P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D. S., ... & Gujral, M. (2017). Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nature genetics, 49(1), 27.

Mattiaccio, L. M., Coman, I. L., Thompson, C. A., Fremont, W. P., Antshel, K. M., & Kates, W. R. (2018). Frontal dysconnectivity in 22q11. 2 deletion syndrome: an atlas-based functional connectivity analysis. Behavioral and Brain Functions, 14(1), 2.

Mas, S., Bernardo, M., Gassó, P., Álvarez, S., Garcia‐Rizo, C., Bioque, M., ... & Lafuente, A. (2010). A functional variant provided further evidence for the association of ARVCF with schizophrenia. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 153(5), 1052-1059.

Mas, S., Bernardo, M., Parellada, E., Garcia-Rizo, C., Gassó, P., Álvarez, S., & Lafuente, A. (2009). ARVCF single marker and haplotypic association with schizophrenia. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 33(6), 1064-1069.

McDonald-McGinn DM, Emanuel BS, Zackai EH. 22q11.2 Deletion Syndrome. In: Pagon RA, Adam MP, Ardinger HH, Wallace SE, Amemiya A, Bean LJ, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle; 1999. http://www.ncbi.nlm.nih.gov/books/NBK1523/.

McDonald-McGinn, D. M., Fahiminiya, S., Revil, T., Nowakowska, B. A., Suhl, J., Bailey, A., ... & Sullivan, K. E. (2012). Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and contribute to atypical findings in patients with 22q11. 2DS. Journal of medical genetics, jmedgenet-2012.

McDonald-McGinn, D. M., Sullivan, K. E., Marino, B., Philip, N., Swillen, A., Vorstman, J. A., ... Bassett, A. S. (2015). 22q11.2 deletion syndrome. Nature Reviews Disease Primers, 1, 1–19.

Meechan, D. W., Maynard, T. M., Tucker, E. S., Fernandez, A., Karpinski, B. A., Rothblat, L. A., & LaMantia, A. S. (2015). Modeling a model: Mouse genetics, 22q11. 2 Deletion Syndrome, and disorders of cortical circuit development. Progress in neurobiology, 130, 1-28.

Meyer, M. J., Das, J., Wang, X., & Yu, H. (2013). INstruct: a database of high-quality 3D structurally resolved protein interactome networks. Bioinformatics, 29(12), 1577-1579.

Meyer-Lindenberg, A., and Weinberger, D. R. (2006). Intermediate phenotypes and genetic mechanisms of psychiatric disorders. Nature Reviews Neuroscience, 7(10), 818.

Merico, D. et al. MicroRNA dysregulation, gene networks, and risk for schizophrenia in 22q11.2 deletion syndrome. Front. Neurol. 5, 238 (2014).

Mignogna ML, Giannandrea M, Gurgone A, Fanelli F, Raimondi F, Mapelli L, S Bassani, Fang H, Van Anken E, Alessio M, Passafaro M, Gatti S, Esteban JA, Huganir R, D'Adamo P. 2015. The intellectual disability protein RAB39B selectively regulates GluA2 trafficking to determine synaptic AMPAR composition. Nat Commun 18; 6:6504.

Mikhail, F. M., Descartes, M., Piotrowski, A., Andersson, R., Diaz de Ståhl, T., Komorowski, J., ... & Carroll, A. J. (2007). A previously unrecognized microdeletion syndrome on chromosome 22 band q11. 2 encompassing the BCR gene. American Journal of Medical Genetics Part A, 143(18), 2178-2184.

Mikhail, F. M., Burnside, R. D., Rush, B., Ibrahim, J., Godshalk, R., Rutledge, S. L., ... & Carroll, A. J. (2014). The recurrent distal 22q11. 2 microdeletions are often de novo and do not represent a single clinical entity: a proposed categorization system. Genetics in medicine, 16(1), 92.

Moghaddam, B., & Javitt, D. (2012). From revolution to evolution: the glutamate hypothesis of schizophrenia and its implication for treatment. Neuropsychopharmacology, 37(1), 4.

Moeschler JB, Shevell M. 2014. Comprehensive evaluation of the child with intellectual disability or global developmental delays. Pediatrics, 134(3): e903-e918.

Molck, M. C., Vieira, T. P., Sgardioli, I. C., Simioni, M., Dos Santos, A. P., Souza, J., ... & Gil-da-Silva-Lopes, V. L. (2013). Atypical copy number abnormalities in 22q11. 2 region: Report of three cases. European journal of medical genetics, 56(9), 515-520.

Molck, M. C., Vieira, T. P., Simioni, M., Sgardioli, I. C., Santos, A. P. D., Xavier, A. C., & Gil‐da‐Silva‐Lopes, V. L. (2015). Distal 22q11. 2 microduplication combined with typical 22q11. 2 proximal deletion: a case report. American Journal of Medical Genetics Part A, 167(1), 215-220.

Mosca, S. J., Langevin, L. M., Dewey, D., Innes, A. M., Lionel, A. C., Marshall, C. C., ... & Bernier, F. P. (2016). Copy-number variations are enriched for neurodevelopmental genes in children with developmental coordination disorder. Journal of medical genetics, jmedgenet-2016.

Muhle, R. A., Reed, H. E., Stratigos, K. A., & Veenstra-VanderWeele, J. (2018). The Emerging Clinical Neuroscience of Autism Spectrum Disorder: A Review. JAMA psychiatry, 75(5), 514-523.

Mukai, J., Liu, H., Burt, R. A., Swor, D. E., Lai, W. S., Karayiorgou, M., & Gogos, J. A. (2004). Evidence that the gene encoding ZDHHC8 contributes to the risk of schizophrenia. Nature genetics, 36(7), 725.

Mukai, J., Tamura, M., Fénelon, K., Rosen, A. M., Spellman, T. J., Kang, R., ... & Gogos, J. A. (2015). Molecular substrates of altered axonal growth and brain connectivity in a mouse model of schizophrenia. Neuron, 86(3), 680-695.

Nacak, T. G., Leptien, K., Fellner, D., Augustin, H. G., & Kroll, J. (2006). The BTB-kelch protein LZTR-1 is a novel Golgi protein that is degraded upon induction of apoptosis. Journal of Biological Chemistry, 281(8), 5065-5071.

Newbern, J., Zhong, J., Wickramasinghe, S. R., Li, X., Wu, Y., Samuels, I., ... & Gargesha, M. (2008). Mouse and human phenotypes indicate a critical conserved role for ERK2 signaling in neural crest development. Proceedings of the National Academy of Sciences.

Nguyen, L. T., Fleishman, R., Flynn, E., Prasad, R., Moulick, A., Mesia, C. I., ... & Jethva, R. (2017). 22q11. 2 microduplication syndrome with associated esophageal atresia/tracheo‐esophageal fistula and vascular ring. Clinical case reports, 5(3), 351-356.

Nik-Zainal, S., Strick, R., Storer, M., Huang, N., Rad, R., Willatt, L., ... & Janecke, A. R. (2011). High incidence of recurrent copy number variants in patients with isolated and syndromic Müllerian aplasia. Journal of medical genetics, jmg-2010.

Niklasson, L., Rasmussen, P., Óskarsdóttir, S., & Gillberg, C. (2009). Autism, ADHD, mental retardation and behavior problems in 100 individuals with 22q11 deletion syndrome. Research in developmental disabilities, 30(4), 763-773.

Nolze, A., Schneider, J., Keil, R., Lederer, M., Hüttelmaier, S., Kessels, M. M., ... & Hatzfeld, M. (2013). FMRP regulates actin filament organization via the armadillo protein p0071. Rna.

Olsen, L., Sparsø, T., Weinsheimer, S. M., Dos Santos, M. B. Q., Mazin, W., Rosengren, A., ... & Bybjerg-Grauholm, J. (2018). Prevalence of rearrangements in the 22q11. 2 region and population-based risk of neuropsychiatric and developmental disorders in a Danish population: a case-cohort study. The Lancet Psychiatry.

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., ... & Astashyn, A. (2015). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic acids research, 44(D1), D733-D745.

O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., ... & Munson, J. (2012). Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science, 338(6114), 1619-1622.

Ottet, M. C., Schaer, M., Cammoun, L., Schneider, M., Debbane, M., Thiran, J. P., & Eliez, S. (2013). Reduced fronto-temporal and limbic connectivity in the 22q11. 2 deletion syndrome: vulnerability markers for developing schizophrenia?. PloS one, 8(3), e58429.

Ou, Z., Berg, J. S., Yonath, H., Enciso, V. B., Miller, D. T., Picker, J., ... & Chinault, A. C. (2008). Microduplications of 22q11. 2 are frequently inherited and are associated with variable phenotypes. Genetics in Medicine, 10(4), 267.

Paronett, E. M., Meechan, D. W., Karpinski, B. A., LaMantia, A. S., & Maynard, T. M. (2014). Ranbp1, deleted in DiGeorge/22q11. 2 deletion syndrome, is a microcephaly gene that selectively disrupts layer 2/3 cortical projection neuron generation. Cerebral cortex, 25(10), 3977-3993.

Philip, N., & Bassett, A. (2011). Cognitive, behavioural and psychiatric phenotype in 22q11. 2 deletion syndrome. Behavior genetics, 41(3), 403-412.

Piccione, M., Vecchio, D., Cavani, S., Malacarne, M., Pierluigi, M., & Corsello, G. (2011). The first case of myoclonic epilepsy in a child with a de novo 22q11. 2 microduplication. American Journal of Medical Genetics Part A, 155(12), 3054-3059.

Pieters, T., Goossens, S., Haenebalcke, L., Andries, V., Stryjewska, A., De Rycke, R., ... & Stemmler, M. P. (2016). p120 Catenin-mediated stabilization of E-cadherin is essential for primitive endoderm specification. PLoS genetics, 12(8), e1006243.

Pinchefsky, E., Laneuville, L., & Srour, M. (2017). Distal 22q11. 2 microduplication: case report and review of the literature. Child neurology open, 4, 2329048X17737651.

Piotrowski, A., Xie, J., Liu, Y. F., Poplawski, A. B., Gomes, A. R., Madanecki, P., ... & Babovic-Vuksanovic, D. (2014). Germline loss-of-function mutations in LZTR1 predispose to an inherited disorder of multiple schwannomas. Nature genetics, 46(2), 182.

Portnoï, M. F., Lebas, F., Gruchy, N., Ardalan, A., Biran‐Mucignat, V., Malan, V., ... & Taillemite, J. L. (2005). 22q11. 2 duplication syndrome: two new familial cases with some overlapping features with DiGeorge/velocardiofacial syndromes. American Journal of Medical Genetics Part A, 137(1), 47-51.

Portnoï, M. F. (2009). Microduplication 22q11. 2: a new chromosomal syndrome. European Journal of Medical Genetics, 52(2), 88-93.

Radoeva, P. D., Coman, I. L., Salazar, C. A., Gentile, K. L., Higgins, A. M., Middleton, F. A., ... & Kates, W. R. (2014). Association between Autism Spectrum Disorder (ASD) in Individuals with Velo-Cardio-Facial (22q11. 2 Deletion) Syndrome and PRODH and COMT Genotypes. Psychiatric genetics, 24(6), 269.

Rafiei, A., Mian, A. A., Döring, C., Metodieva, A., Oancea, C., Thalheimer, F. B., ... & Ruthardt, M. (2015). The functional interplay between the t (9; 22)-associated fusion proteins BCR/ABL and ABL/BCR in Philadelphia chromosome-positive acute lymphatic leukemia. PLoS genetics, 11(4), e1005144.

Rauch, A., Pfeiffer, R. A., Leipold, G., Singer, H., Tigges, M., & Hofbeck, M. (1999). A novel 22q11. 2 microdeletion in DiGeorge syndrome. American journal of human genetics, 64(2), 659.

Rauch, A., Zink, S., Zweier, C., Thiel, C. T., Koch, A., Rauch, R., ... & Hofbeck, M. (2005). Systematic assessment of atypical deletions reveals genotype–phenotype correlation in 22q11. 2. Journal of medical genetics, 42(11), 871-876.

Ribeiro-Bicudo, L. A., de Campos Legnaro, C., Gamba, B. F., Sandri, R. C., & Richieri-Costa, A. (2013). Cognitive deficit, learning difficulties, severe behavioral abnormalities and healed cleft lip in a patient with a 1.2-Mb distal microduplication at 22q11. 2. Molecular syndromology, 4(6), 292-296.

Rødningen, O. K., Prescott, T., Eriksson, A. S., & Røsby, O. (2008). 1.4 Mb recurrent 22q11. 2 distal deletion syndrome, two new cases expand the phenotype. European journal of medical genetics, 51(6), 646-650.

Rodriguez-Santiago, B., Brunet, A., Sobrino, B., Serra-Juhe, C., Flores, R., Armengol, L., ... & Martorell, L. (2010). Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. Molecular psychiatry, 15(10), 1023.

Saitta, S. C., McGrath, J. M., Mensch, H., Shaikh, T. H., Zackai, E. H., & Emanuel, B. S. (1999). A 22q11. 2 deletion that excludes UFD1L and CDC45L in a patient with conotruncal and craniofacial defects. American journal of human genetics, 65(2), 562.

Sanders, A. R., Rusu, I., Duan, J., Vander Molen, J. E., Hou, C., Schwab, S. G., ... & Gejman, P. V. (2005). Haplotypic association spanning the 22q11. 21 genes COMT and ARVCF with schizophrenia. Molecular psychiatry, 10(4), 353.

Sanders, S. J., He, X., Willsey, A. J., Ercan-Sencicek, A. G., Samocha, K. E., Cicek, A. E., ... & Goldberg, A. P. (2015). Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron, 87(6), 1215-1233.

Schneider, M., Debbané, M., Bassett, A. S., Chow, E. W., Fung, W. L. A., Van Den Bree, M. B., ... & Antshel, K. M. (2014). Psychiatric disorders from childhood to adulthood in 22q11. 2 deletion syndrome: results from the International Consortium on Brain and Behavior in 22q11. 2 Deletion Syndrome. American Journal of Psychiatry, 171(6), 627-639.

Sedghi, M., Abdali, H., Memarzadeh, M., Salehi, M., Nouri, N., Hosseinzadeh, M., & Nouri, N. (2015). Identification of proximal and distal 22q11. 2 microduplications among patients with cleft lip and/or palate: a novel inherited atypical 0.6 Mb duplication. Genetics research international, 2015.

Seong, E., Yuan, L., & Arikkath, J. (2015). Cadherins and catenins in dendrite and synapse morphogenesis. Cell adhesion & migration, 9(3), 202-213.

Sgardioli, I. C., de Mello Copelli, M., Monteiro, F. P., dos Santos, A. P., Mendes, E. L., Vieira, T. P., & Gil-da-Silva-Lopes, V. L. (2017). Diagnostic Approach to Microdeletion

Syndromes Based on 22q11. 2 Investigation: Challenges in Four Cases. Molecular syndromology, 8(5), 244-252.

Shaikh, T. H., O'Connor, R. J., Pierpont, M. E., McGrath, J., Hacker, A. M., Nimmakayalu, M., ... & Saitta, S. C. (2007). Low copy repeats mediate distal chromosome 22q11. 2 deletions: sequence analysis predicts breakpoint mechanisms. Genome research, 17(4), 000-000.

Shi, H., & Wang, Z. (2018). Atypical microdeletion in 22q11 deletion syndrome reveals new candidate causative genes: A case report and literature review. Medicine, 97(8).

Shimojima, K., Imai, K., & Yamamoto, T. (2010). A de novo 22q11. 22q11. 23 interchromosomal tandem duplication in a boy with developmental delay, hyperactivity, and epilepsy. American Journal of Medical Genetics Part A, 152(11), 2820-2826.

Shishido, E., Aleksic, B., & Ozaki, N. (2014). Copy‐number variation in the pathogenesis of autism spectrum disorder. Psychiatry and clinical neurosciences, 68(2), 85-95.

Sholl, D. A. (1950). The theory of differential growth analysis. Proc. R. Soc. Lond. B, 137(889), 470-474.

Sirotkin, H., O'Donnell, H., DasGupta, R., Halford, S., Jore, B. S., Puech, A., ... & Scambler, P. (1997). Identification of a New Human Catenin Gene Family Member (ARVCF) from the Region Deleted in Velo–Cardio–Facial Syndrome. Genomics, 41(1), 75-83.

Spineli-Silva, S., Bispo, L. M., Gil-da-Silva-Lopes, V. L., & Vieira, T. P. (2018). Distal deletion at 22q11. 2 as differential diagnosis in Craniofacial Microsomia: Case report and literature review. European journal of medical genetics, 61(5), 262-268.

Spletter, M. L., Liu, J., Liu, J., Su, H., Giniger, E., Komiyama, T., ... & Luo, L. (2007). Lola regulates Drosophila olfactory projection neuron identity and targeting specificity. Neural development, 2(1), 14.

Stark, K. L. et al. Altered brain microRNA biogenesis contributes to phenotypic deficits in a 22q11-deletion mouse model. Nat. Genet. 40, 751–760 (2008).

Stessman, H. A., Bernier, R., & Eichler, E. E. (2014). A genotype-first approach to defining the subtypes of a complex disease. Cell, 156(5), 872-877.

Stessman, H. A., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., ... & Vives, L. (2017). Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. Nature genetics, 49(4), 515.

Stogios, P. J., Downs, G. S., Jauhal, J. J., Nandra, S. K., & Privé, G. G. (2005). Sequence and structural analysis of BTB domain proteins. Genome biology, 6(10), R82.

Stoll, G., Pietiläinen, O. P., Linder, B., Suvisaari, J., Brosi, C., Hennah, W., ... & Plöttner, O. (2013). Deletion of TOP3β, a component of FMRP-containing mRNPs, contributes to neurodevelopmental disorders. Nature neuroscience, 16(9), 1228.

Sullivan, P. F., Daly, M. J., & O'donovan, M. (2012). Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nature Reviews Genetics, 13(8), 537.

Sun, D., Ching, C. R. K., Lin, A., Forsyth, J. K., Kushan, L., Vajdi, A., ... & Jonas, R. K. (2018). Large-scale mapping of cortical alterations in 22q11. 2 deletion syndrome: Convergence with idiopathic psychosis and effects of deletion size. Molecular psychiatry.

Szatkiewicz, J. P., O'Dushlaine, C., Chen, G., Chambert, K., Moran, J. L., Neale, B. M., ... & Kähler, A. (2014). Copy number variation in schizophrenia in Sweden. Molecular psychiatry, 19(7), 762.

Swillen, A., & McDonald‐McGinn, D. (2015, June). Developmental trajectories in 22q11. 2 deletion syndrome. In American Journal of Medical Genetics Part C: Seminars in Medical Genetics (Vol. 169, No. 2, pp. 172-181).

Tan, T. Y., Collins, A., James, P. A., McGillivray, G., Stark, Z., Gordon, C. T., ... & Ganesamoorthy, D. (2011). Phenotypic variability of distal 22q11. 2 copy number abnormalities. American Journal of Medical Genetics Part A, 155(7), 1623-1633.

Thompson, C. A., Karelis, J., Middleton, F. A., Gentile, K., Coman, I. L., Radoeva, P. D., ... & Kates, W. R. (2017). Associations between neurodevelopmental genes, neuroanatomy, and ultra high risk symptoms of psychosis in 22q11. 2 deletion syndrome. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 174(3), 295-314.

Toritsuka, M., Kimoto, S., Muraki, K., Landek-Salgado, M. A., Yoshida, A., Yamamoto, N., ... & Illingworth, E. (2013). Deficits in microRNA-mediated Cxcr4/Cxcl12 signaling in neurodevelopmental deficits in a 22q11 deletion syndrome mouse model. Proceedings of the National Academy of Sciences, 110(43), 17552-17557.

Torti, E. E., Braddock, S. R., Bernreuter, K., & Batanian, J. R. (2013). Oculo‐auriculo‐vertebral spectrum, cat eye, and distal 22q11 microdeletion syndromes: A unique double rearrangement. American Journal of Medical Genetics Part A, 161(8), 1992-1998.

Toth, G., Zraly, C. B., Thomson, T. L., Jones, C., Lapetino, S., Muraskas, J., ... & Dingwall, A. K. (2011). Congenital anomalies and rhabdoid tumor associated with 22q11 germline deletion and somatic inactivation of the SMARCB1 tumor suppressor. Genes, Chromosomes and Cancer, 50(6), 379-388.

Tunbridge, E., Burnet, P. W., Sodhi, M. S., & Harrison, P. J. (2004). Catechol‐o‐methyltransferase (COMT) and proline dehydrogenase (PRODH) mRNAs in the dorsolateral prefrontal cortex in schizophrenia, bipolar disorder, and major depression. Synapse, 51(2), 112-118.

Tunbridge, E. M., Harrison, P. J., & Weinberger, D. R. (2006). Catechol-o-methyltransferase, cognition, and psychosis: Val158Met and beyond. Biological psychiatry, 60(2), 141-151.

Turner, T. N., Sharma, K., Oh, E. C., Liu, Y. P., Collins, R. L., Sosa, M. X., ... & Pihur, V. (2015). Loss of δ-catenin function in severe autism. Nature, 520(7545), 51.

Turner, T. N., Yi, Q., Krumm, N., Huddleston, J., Hoekzema, K., F. Stessman, H. A., ... & Eichler, E. E. (2016). denovo-db: a compendium of human de novo variants. Nucleic acids research, 45(D1), D804-D811.

Van Campenhout, S., Devriendt, K., Breckpot, J., Fryns, J. P., Peeters, H., Van Buggenhout, G., ... & Swillen, A. (2012). Microduplication 22q11. 2: a description of the clinical, developmental and behavioral characteristics during childhood. Genetic Counseling, 23(2), 135-148.

Van, L., Boot, E., & Bassett, A. S. (2017). Update on the 22q11. 2 deletion syndrome and its relevance to schizophrenia. Current opinion in psychiatry, 30(3), 191-196.

van Karnebeek CD, Shevell M, Zschocke J, Moeschler JB, Stockler S. 2014. The metabolic evaluation of the child with an intellectual developmental disorder: diagnostic algorithm for identification of treatable causes and new digital resource. Molecular genetics and metabolism, 111(4): 428-438.

Vecchio D, Piccione M, D'Adamo P, Mignogna M, Salzano E, Giuffrè M, Antona V, Caputo V, Pizzuti A, Nardello R, Piro E, Capobianco E, Corsello G. 2016. Intellectual disability, epilepsy and mild dysmorphisms due 22q11.2 distal duplication: clinical and

molecular characterization of a 0.5 mb minimal critical region. Eur J Pediatr. Part of DOI:10.1007/s00431-016-2785-8

Velasquez, K. M., Molfese, D. L., & Salas, R. (2014). The role of the habenula in drug addiction. Frontiers in human neuroscience, 8, 174.

Verhoeven, W., Egger, J., Brunner, H., & de Leeuw, N. (2011). A patient with a de novo distal 22q11. 2 microdeletion and anxiety disorder. American Journal of Medical Genetics Part A, 155(2), 392-397.

Vorstman, J. A., Turetsky, B. I., Sijmens-Morcus, M. E., De Sain, M. G., Dorland, B., Sprong, M., ... & Van Engeland, H. (2009a). Proline affects brain function in 22q11DS children with the low activity COMT 158 allele. Neuropsychopharmacology, 34(3), 739.

Vorstman, J. A., Chow, E. W., Ophoff, R. A., van Engeland, H., Beemer, F. A., Kahn, R. S., ... & Bassett, A. S. (2009b). Association of the PIK4CA schizophrenia‐susceptibility gene in adults with the 22q11. 2 deletion syndrome. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics, 150(3), 430-433.

Vorstman, J. A., Breetvelt, E. J., Thode, K. I., Chow, E. W., & Bassett, A. S. (2013). Expression of autism spectrum and schizophrenia in patients with a 22q11. 2 deletion. Schizophrenia research, 143(1), 55-59.

Vorstman, J. A., Breetvelt, E. J., Duijff, S. N., Eliez, S., Schneider, M., Jalbrzikowski, M., ... & Chow, E. W. (2015). Cognitive decline preceding the onset of psychosis in patients with 22q11. 2 deletion syndrome. JAMA psychiatry, 72(4), 377-385.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., & Blelloch, R. (2007). DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. Nature genetics, 39(3), 380.

Ware, J. S., Samocha, K. E., Homsy, J., & Daly, M. J. (2015). Interpreting de novo variation in human disease using denovolyzeR. Current protocols in human genetics, 87(1), 7-25.

Wieser, R., Fritz, B., Ullmann, R., Müller, I., Galhuber, M., Storlazzi, C. T., ... & Rehder, H. (2005). Novel rearrangement of chromosome band 22q11. 2 causing 22q11 microdeletion syndrome‐like phenotype and rhabdoid tumor of the kidney. Human mutation, 26(2), 78-83.

Williams, N. M., Franke, B., Mick, E., Anney, R. J., Freitag, C. M., Gill, M., ... & Kent, L. (2012). Genome-wide analysis of copy number variants in attention deficit

hyperactivity disorder: the role of rare variants and duplications at 15q13. 3. American Journal of Psychiatry, 169(2), 195-204.

Wincent, J., Bruno, D. L., Van Bon, B. W. M., Bremer, A., Stewart, H., Bongers, E. M. H. F., ... & Newbury, D. F. (2010). Sixteen new cases contributing to the characterization of patients with distal 22q11. 2 microduplications. Molecular syndromology, 1(5), 246-254.

Xu, J., Fan, Y. S., & Siu, V. M. (2008). A child with features of Goldenhar syndrome and a novel 1.12 Mb deletion in 22q11. 2 by cytogenetics and oligonucleotide array CGH: is this a candidate region for the syndrome?. American Journal of Medical Genetics Part A, 146(14), 1886-1889.

Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A., & Dougherty, J. D. (2014). Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. Journal of Neuroscience, 34(4), 1420-1431.

Yamamoto, G. L., Aguena, M., Gos, M., Hung, C., Pilch, J., Fahiminiya, S., ... & Malaquias, A. C. (2015). Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. Journal of medical genetics, 52(6), 413-421.

Yu, S., Graf, W. D., Ramalingam, A., Brawner, S. J., Joyce, J. M., Fiedler, S., ... & Liu, H. Y. (2011). Identification of copy number variants on human chromosome 22 in patients with a variety of clinical findings. Cytogenetic and genome research, 134(4), 260-268.

Yuan, L., & Arikkath, J. (2017, September). Functional roles of p120ctn family of proteins in central neurons. In Seminars in cell & developmental biology (Vol. 69, pp. 70-82). Academic Press.

Zebda, N., Tian, Y., Tian, X., Gawlak, G., Higginbotham, K., Reynolds, A. B., ... & Birukov, K. G. (2013). Interaction of p190RhoGAP with C-terminal domain of p120-catenin modulates endothelial cytoskeleton and permeability. Journal of Biological Chemistry, jbc-M112.

Zeng, Z. L., Li, F. J., Gao, F., Sun, D. S., & Yao, L. (2013). Upregulation of miR‐650 is correlated with down‐regulation of ING4 and progression of hepatocellular carcinoma. Journal of surgical oncology, 107(2), 105-110.

Zhou, J. Y., Fogelgren, B., Wang, Z., Roe, B. A., & Biegel, J. A. (2000). Isolation of genes from the rhabdoid tumor deletion region in chromosome band 22q11. 2. Gene, 241(1), 133-141.

ANNEXES

Supplementary Information and Notes

ExAC (brief description): Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a variety of large-scale sequencing projects, and to make summary data available for the wider scientific community. The genetic information provided spans 60,706 unrelated individuals. Variants identified in patients affected by severe pediatric diseases were removed. Thus, these data can serve as a useful reference set of allele frequencies. All data are released publicly for the benefit of the wider biomedical community and, are available under the ODC Open Database License (ODbL).

denovo-db (brief description): As of July 2016, denovo-db contained 40 different studies and 32,991 de novo variants from 23,098 trios. Database features include basic variant information (chromosome location, change, type); detailed annotation at the transcript and protein levels; severity scores; frequency; validation status; and, most importantly, the phenotype of the individual with the variant. denovo-db provides necessary information for researchers to compare their data to other individuals with the same phenotype and also to controls allowing for a better understanding of the biology of de novo variants and their contribution to disease. All data are released publicly for the benefit of the wider biomedical community and, are available under the ODC Open Database License (ODbL).

CNV morbidity map: Coe BP et al., 2014 constructed an expanded CNV morbidity map using array comparative genomic hybridization (CGH) data from 29,085 primarily pediatric patients with intellectual disability, developmental delay and/or ASD in comparison to 19,584 adult population controls. The set included 13,318 previously unpublished cases and 11,255 controls, providing enhanced power to detect large-scale, potentially pathogenic deletions and duplications. A total of 29,415 rare autosomal CNVs in cases and 741,729 (289,359 new) control CNVs were detected and deposited into dbVar (study accession nstd100). CNV burden was compared between cases and controls for rare CNVs (frequency of <1%) using CNV length excluding gaps and regions annotated as segmental duplications (hg18). Burden was defined using only the largest CNV to account for the large number of bases encompassed by small CNVs and the difference in array resolution between cases and controls. Statistical comparisons used the Peto and Peto modification of the Gehan- Wilcoxon test (because of non-proportional hazard ratios) to assess overall burden. For significance at specific thresholds, the authors used the Fisher's exact test. Significance for CNV enrichment was enumerated for all RefSeq genes (NCBI Build 36). All isoforms for each gene were combined into a single entry representing all possible coding bases. Rare CNVs from cases and all control CNVs were then enumerated for only cases where the CNV intersected with an exon. The resulting counts were compared using the one-tailed

Fisher's exact test. Additionally, the authors also calculated an empirical p-value for genes affected by rare CNVs. To do so, they first excluded CNVs residing in regions with elevated mutation rates or unreliable CNV detection. These regions included subtelomeric CNVs initiating in the first 1.5 Mb of each chromosome, over 75% of bases intersecting with hotspots (145.1 Mb across 58 sites) and segmental duplications (130.4 Mb across 7,264 sites), initiating or terminating in a centromere gap region. All CNVs under 10 Mb in length were then randomly shuffled (chromosome selection was weighted by the number of bases not filtered) under these constraints for cases and controls, and Fisher's exact tests were calculated 20,000 times for deletions and duplications of each gene. The empirical p-values, also used in our project, is therefore defined as the number of simulations more significant than observed plus one divided by the number of simulations plus one.

Model 1 formula: The assumes an underlying binomial distribution for the *de novo* mutations (i.e., each event is a Bernoulli trial). The probability of success (i.e., *de novo* mutability) incorporates the overall mutation rate using human chimpanzee fixed differences in coding sequence and splice sites. Protein-altering events are weighted (w) across protein-encoding transcriptome (18,946 genes; labelled with subscript i):

$$w_{protein-altering,i} = w_{missense,i} + w_{truncating,i} = w_{missense,i} + (w_{nonsense,i} + w_{splice,i} + w_{indel,i})$$

The protein-altering weight is then normalized across all genes:

$$\sum_{i=1}^{18,946} (w_{missense,i} + w_{truncating,i}) = 1.$$

The normalized weights per gene are converted into probability of success (p), to estimate overall probability of at least the observed count of *de novo* mutations per gene i (m_i) using the following binomial framework:

$$P(X \geq m_i) = 1 - \sum_{x_i=0}^{m_i} \left[\binom{n}{x_i} \times p^{x_i} \times (1-p)^{(n-x_i)} \right]$$

where n is the total number of *de novo* SNVs. As $n \to \infty, p \to 0$ such that $n \times p = \lambda$. The two models show generally very good agreement (r = 0.9) but differences do exist due to evolutionary idiosyncrasies since human–chimpanzee divergence within the human lineage.

Model 2 formula: In this model, the probability of each nucleotide mutating into one of the other three bases is calculated in the context of 192 trinucleotides. Briefly, mutability of the central base in a trinucleotide, e.g., ACA (chimpanzee) → AAA (human), is measured as the proportion of mutations affecting that base divided by the total count of the trinucleotide in the human genome. The end product is a probability

table for 192 possible trinucleotide divergence events. These probabilities are then added either across the entire gene (producing gene-level probability) or each annotation class (producing annotation class-specific mutability probabilities). A Poisson model is used to estimate probability of observing at least *m de novo* mutations in gene *i*, given an expected *de novo* number per-gene (m_i), estimated as sample size × mutability-probability:

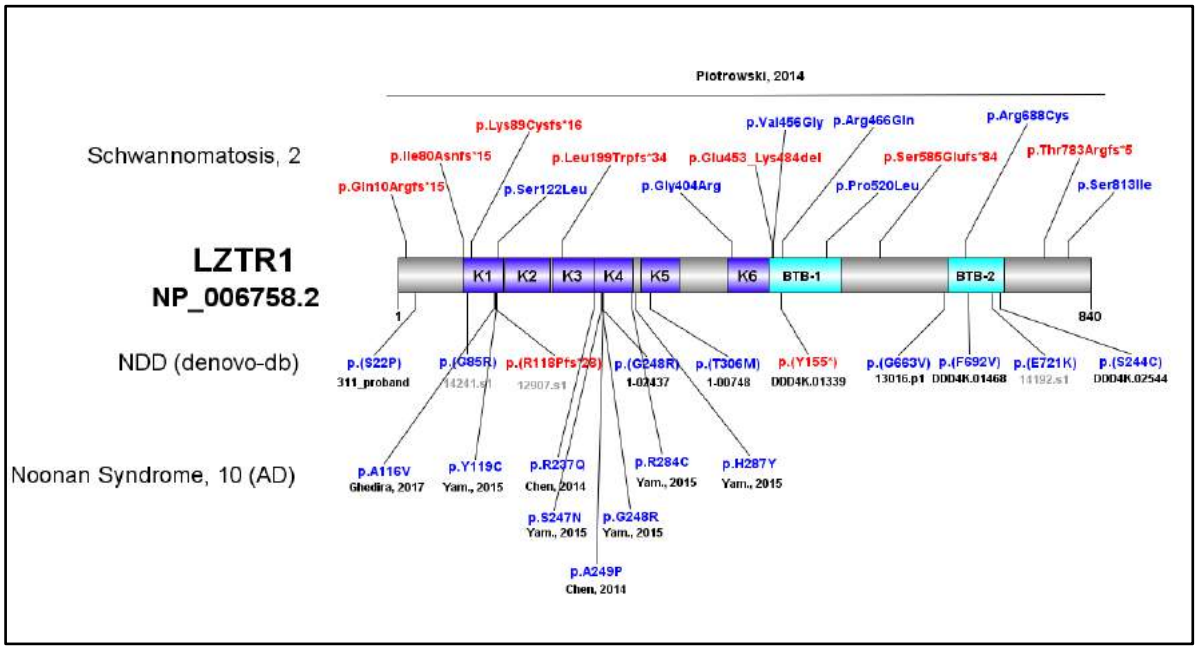
$$P(X \geq m_i) = 1 - \sum_{x_i=0}^{m_i} \left[\frac{\lambda^{x_i} \times e^{-\lambda}}{x_i!} \right]$$

Web Resources

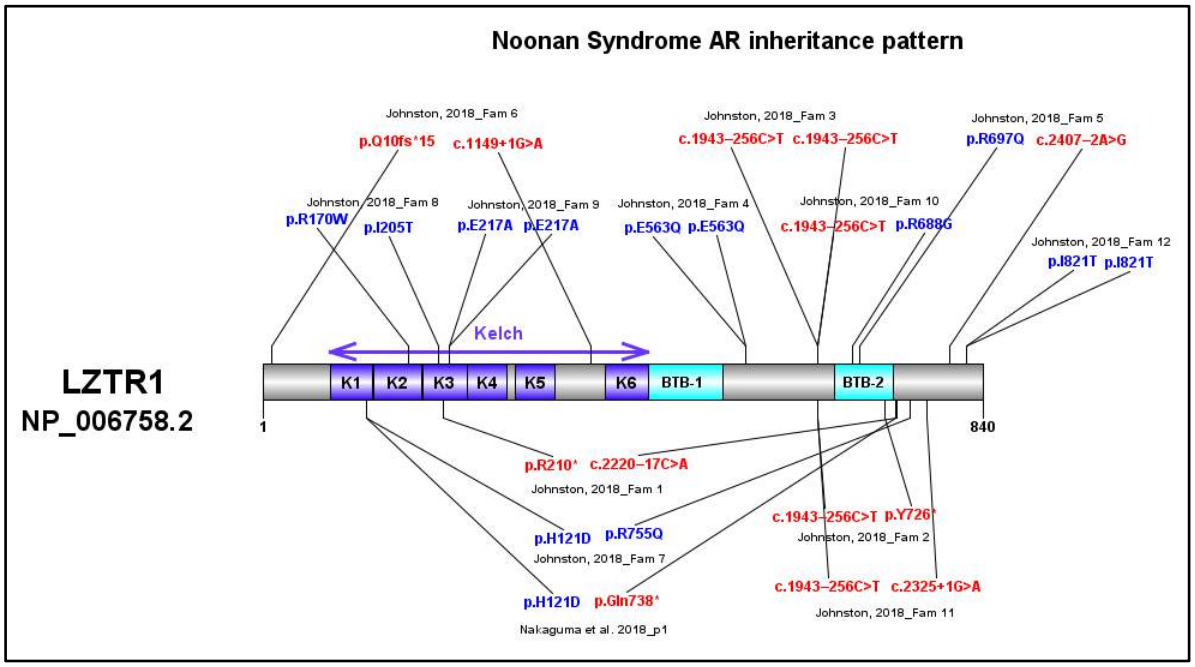
The URLs for data herein presented are as follows:

- Gene Expression Omnibus (GEO): http://www.ncbi.nlm.nih.gov/geo/
- Online Mendelian Inheritance in Man (OMIM): http://www.omim.org/
- RefSeq: http://www.ncbi.nlm.nih.gov/RefSeq
- UCSC Genome Browser: http://genome.ucsc.edu
- GTExPortal: https://gtexportal.org/home/
- CSEA tool: http://genetics.wustl.edu/jdlab/csea-tool-2/
- ToppGenes Suite: https://toppgene.cchmc.org
- Denovo-db, http://denovo-db.gs.washington.edu/denovo- db/;
- Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org/;
- dbVar: https://www.ncbi.nlm.nih.gov/dbvar
- denovolyzeR: http://denovolyzer.org
- CADD: https://cadd.gs.washington.edu
- MIPgen: http://shendurelab.github.io/MIPGEN
- dbSNP: https://www.ncbi.nlm.nih.gov/projects/SNP
- SeattleSeq138: http://snp.gs.washington.edu/SeattleSeqAnnotation138
- Ensembl Variant Effect Predictor: https://www.ensembl.org/info/docs/tools/vep
- SFARI: https://gene.sfari.org
- Charité HPO browser: http://compbio.charite.de/hpoweb/
- BrainSpan: http://www.brainspan.org
- Instruct: http://instruct.yulab.org
- MuPIT: http://mupit.icm.jhu.edu/MuPIT_Interactive

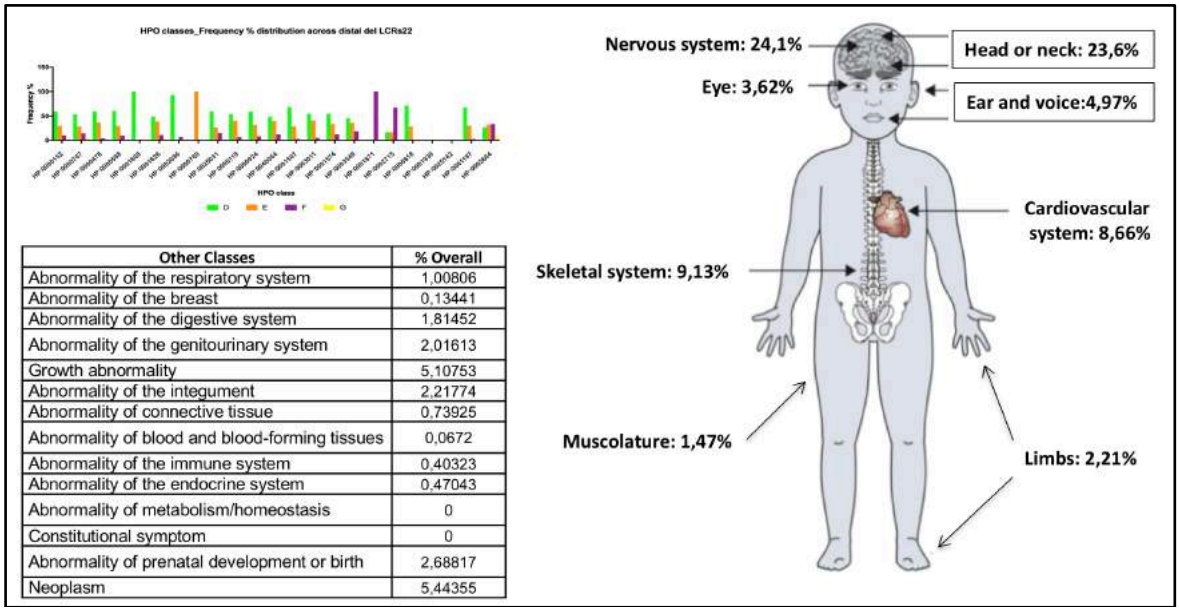
Supplementary Figures



Supplementary Figure 1: Protein diagram of LZRT1 depicting the location of both LGD (red) and missense (blu) previously reported deleterious variants in denovo-db (annotated below the protein structure) compared to those retrieved in patients affected by Noonan Syndrome, 10 and Schwannomatosis, 2. Domain abbreviations: K, kelch; BTB, broad-complex - tramtrack and bric a brac.

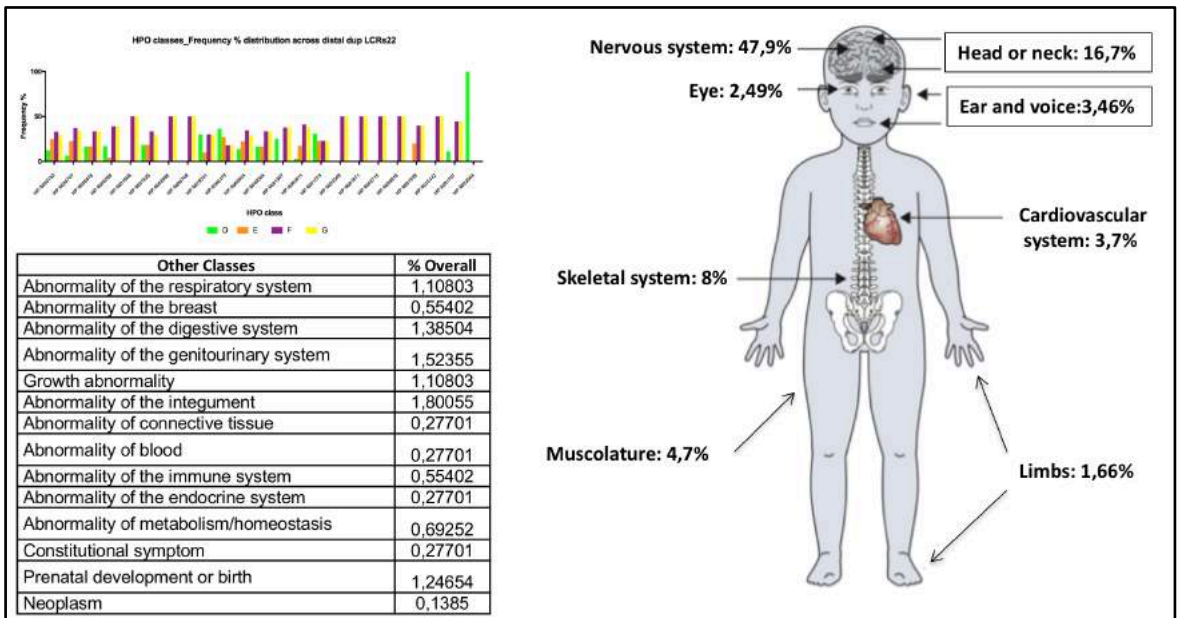


Supplementary Figure 2: Protein diagram of LZRT1 depicting the location of both LGD (red) and missense (blu) previously reported deleterious variants retrieved in patients affected by an autosomal recessive form of Noonan Syndrome. Domain abbreviations: K, kelch; BTB, broad-complex - tramtrack and bric a brac.



Other Classes	% Overall
Abnormality of the respiratory system	1,00806
Abnormality of the breast	0,13441
Abnormality of the digestive system	1,81452
Abnormality of the genitourinary system	2,01613
Growth abnormality	5,10753
Abnormality of the integument	2,21774
Abnormality of connective tissue	0,73925
Abnormality of blood and blood-forming tissues	0,0672
Abnormality of the immune system	0,40323
Abnormality of the endocrine system	0,47043
Abnormality of metabolism/homeostasis	0
Constitutional symptom	0
Abnormality of prenatal development or birth	2,68817
Neoplasm	5,44355

Supplementary Figure 3: Ideogrammatic representation of the congenital anomalies retrieved in 100 available clinical reports of patients carrying a distal 22q11.2 deletion event, reporting the impinged systems and/or apparatuses' percentage frequency distribution.



Other Classes	% Overall
Abnormality of the respiratory system	1,10803
Abnormality of the breast	0,55402
Abnormality of the digestive system	1,38504
Abnormality of the genitourinary system	1,52355
Growth abnormality	1,10803
Abnormality of the integument	1,80055
Abnormality of connective tissue	0,27701
Abnormality of blood	0,27701
Abnormality of the immune system	0,55402
Abnormality of the endocrine system	0,27701
Abnormality of metabolism/homeostasis	0,69252
Constitutional symptom	0,27701
Prenatal development or birth	1,24654
Neoplasm	0,1385

Supplementary Figure 4: Ideogrammatic representation of the congenital anomalies retrieved in 48 available clinical reports of patients carrying a distal 22q11.2 duplication event, reporting the impinged systems and/or apparatuses' percentage frequency distribution.

Supplementary Tables

Supplementary Table 1: 22q11.2 genes' highest median brain GTEx TPM entries, their LCRs22 mapping and first quartile distribution.

GENE	LCR	GTEx Brain Median TPM	GTEx Significative Brain TPM	GTEx 1st quartile
ARVCF	A	175	YES	YES
C22orf29	A	18	YES	NO
C22orf39	A	43	YES	NO
CLDN5	A	76	YES	YES
CLTCL1	A	5	YES	NO
COMT	A	61	YES	YES
DGCR10	A	42	YES	NO
DGCR11	A	3	YES	NO
DGCR14	A	17	YES	NO
DGCR2	A	87	YES	YES
DGCR5	A	48	YES	YES
DGCR6	A	36	YES	NO
DGCR8	A	82	YES	YES
DGCR9	A	63	YES	YES
FAM230A	A	0	NO	NO
GGT3P	A	0	NO	NO
GNB1L	A	2	YES	NO
GSC2	A	0	NO	NO
HIRA	A	43	YES	NO
LINC00896	A	0	NO	NO
MRPL40	A	26	YES	NO
PI4KAP1	A	12	YES	NO
PRODH	A	75	YES	YES
RANBP1	A	56	YES	YES
RIMBP3	A	0	NO	NO
SEPT5	A	359	YES	YES
SLC25A1	A	69	YES	YES
TANGO2	A	14	YES	NO
TBX1	A	0	NO	NO
TRMT2A	A	49	YES	YES
TSSK2	A	0	NO	NO
TXNRD2	A	36	YES	NO
UFD1L	A	31	YES	NO
USP18	A	2	YES	NO
ZDHHC8	A	54	YES	YES
DGCR6L	B	77	YES	YES
KLHL22	B	41	YES	NO
MED15	B	50	YES	YES
MIR1286	B	0	NO	NO
RTN4R	B	91	YES	YES
SCARF2	B	5	YES	NO
USP41	B	0	NO	NO
ZNF74	B	15	YES	NO
AIFM3	C	73	YES	YES
CRKL	C	71	YES	YES
LZTR1	C	63	YES	YES
MIR649	C	0	NO	NO
P2RX6	C	11	YES	NO
P2RX6P	C	10	YES	NO
PI4KA	C	124	YES	YES
POM121L4P	C	0	NO	NO
SERPIND1	C	0	NO	NO
SLC7A4	C	10	YES	NO
SNAP29	C	19	YES	NO
THAP7	C	54	YES	YES
THAP7-AS1	C	5	YES	NO
TMEM191A	C	6	YES	NO
TUBA3FP	C	8	YES	NO
BCRP2	D	0	NO	NO
CCDC116	D	0	NO	NO
FAM230B	D	0	NO	NO
GGT2	D	0	NO	NO
HIC2	D	6	YES	NO
MAPK1	D	65	YES	YES
MIR130B	D	0	NO	NO
MIR301B	D	0	NO	NO
PI4KAP2	D	21	YES	NO
POM121L8P	D	0	NO	NO
PPIL2	D	30	YES	NO
PRAME	D	0	NO	NO
PRAMENP	D	0	NO	NO
RIMBP3B	D	0	NO	NO
RIMBP3C	D	0	NO	NO
SDF2L1	D	24	YES	NO
TMEM191C	D	2	YES	NO
TOP3B	D	19	YES	NO
UBE2L3	D	86	YES	YES
VPREB1	D	0	NO	NO
YDJC	D	30	YES	NO
YPEL1	D	22	YES	NO
ZNF280B	D	6	YES	NO
BCR	E	54	YES	YES
FBXW4P1	E	1	YES	NO
GGTLC2	E	0	NO	NO
GNAZ	E	65	YES	YES
IGLL5	E	0	NO	NO
MIR650	E	0	NO	NO
POM121L1P	E	0	NO	NO
RAB36	E	11	YES	NO
RTDR1	E	3	YES	NO
C22orf15	F	0	NO	NO
CABIN1	F	75	YES	YES
CES5AP1	F	3	YES	NO
CHCHD10	F	82	YES	YES
DDT	F	28	YES	NO
DDTL	F	4	YES	NO
DERL3	F	0	NO	NO
GGT5	F	23	YES	NO
GSTT1	F	62	YES	YES
GSTT2	F	6	YES	NO
GSTT2B	F	1	YES	NO
GSTTP1	F	0	NO	NO
IGLL1	F	0	NO	NO
MIF	F	22	YES	NO
MMP11	F	0	NO	NO
RGL4	F	0	NO	NO
SLC2A11	F	40	YES	NO
SMARCB1	F	100	YES	YES
SUSD2	F	0	NO	NO
VPREB3	F	0	NO	NO
ZDHHC8P1	F	23	YES	NO
ZNF70	F	3	YES	NO
ADORA2A	G	46	YES	YES
ADORA2A-AS1	G	2	YES	NO
BCRP3	G	7	YES	NO
GGT1	G	10	YES	NO
GUCD1	G	27	YES	NO
POM121L10P	G	0	NO	NO
POM121L9P	G	0	NO	NO
SNRPD3	G	55	YES	YES
SPECC1L	G	44	YES	YES
UPB1	G	0	NO	NO

Supplementary Table 2: A) Chromosome 22 genomic coordinates of LCR22-A to -H according to human genome NCBI/hg18 and GRCh37/hg19 assemblies and, as reported in Mikhail FM et al., 2014. B) A to G genomic conventional intervals defined between flanking LCRs22. Genomic boundary coordinates are hereby listed according the human GRCh38/hg38 assembly.

A)

LCR22	NCBI36/hg18	GRCh37/hg19
A	~17,020,000–17,290,000	~18,640,000–18,910,000
B	~18,630,000–19,060,000	~20,250,000–20,680,000
C	~19,350,000–19,420,000	~21,020,000–21,090,000
D	~19,800,000–20,250,000	~21,470,000–21,920,000
E	~21,290,000–21,380,000	~22,960,000–23,050,000
F	~21,980,000–22,150,000	~23,650,000–23,820,000
G	~22,960,000–23,030,000	~24,630,000–24,700,000
H	~23,325,000–23,410,000	~24,995,000–25,080,000

B)

22q11.2 genomic conventional intervals	GRCh38/hg38 coordinates
A	18,157,233-20,241,414
B	20,241,415-20,665,711
C	20,665,712-21,115,710
D	21,115,711-22,617,529
E	22,617,530-23,307,812
F	23,307,813-24,234,031
G	24,234,032-24,684,033

Supplementary Table 3: GO cellular component analysis results of the tested 22q11.2 first quartile brain-expressed genes' set; p-values generated by using the ToppGenes Suite tool algorithm. Abbreviation as follow: FDR, false discovery rate; B&H, Benjamini-Hochberg correction; B&Y, Benjamini-Yekutieli correction.

ID	NAME	P-VALUE	FDR B&H	FDR B&Y	BONFERRONI
GO:0030424	axon	1.365E-5	2.429E-3	1.400E-2	2.429E-3
GO:0033267	axon part	7.843E-5	6.980E-3	4.022E-2	1.396E-2
GO:0043005	neuron projection	2.328E-4	1.098E-2	6.324E-2	4.143E-2
GO:0097458	neuron part	2.467E-4	1.098E-2	6.324E-2	4.390E-2
GO:0044463	cell projection part	1.042E-3	3.648E-2	2.102E-1	1.855E-1
GO:0099568	cytoplasmic region	1.230E-3	3.648E-2	2.102E-1	2.189E-1

Supplementary Table 4: Counts of missense and LGD deleterious coding variants seen in the 22q11.2 brain expressed genes retrieved from EXAc database, and broken out *per* observed and expected events.

GENE	22q11.2 Brain expressed 1st quartile	Missense_expected	Missense_obsverd	LGD_expected	LGD_observed
ARVCF	YES	447	473	25	14
C22orf29	NO	128	131	8	8
C22orf39	NO	56	48	5	8
CLDN5	YES	133	53	4	0
CLTCL1	NO	502	558	48	43
COMT	YES	110	100	6	5
DGCR10	YES	0	0	0	0
DGCR11	NO	0	0	0	0
DGCR14	NO	199	182	15	10
DGCR2	YES	120	126	14	7
DGCR5	YES	0	0	0	0
DGCR6	NO	40	45	82	89
DGCR8	YES	322	190	26	0
DGCR9	YES	0	0	0	0
GNB1L	NO	166	156	10	6
HIRA	NO	389	228	41	1
MRPL40	NO	64	72	9	4
PI4KAP1	NO	0	0	0	0
PRODH	YES	189	167	17	11
RANBP1	YES	72	38	8	2
SEPT5	YES	166	68	16	2
SLC25A1	YES	122	80	12	2
TANGO2	NO	101	106	14	10
TRMT2A	YES	260	279	20	11
TXNRD2	NO	213	204	23	16
UFD1L	NO	112	53	17	0
USP18	NO	108	57	13	3
ZDHHC8	YES	329	247	22	2
DGCR6L	YES	86	82	9	3
KLHL22	NO	265	172	17	3
MED15	YES	299	211	36	6
RTN4R	YES	216	172	6	1
SCARF2	NO	0	0	0	0
ZNF74	NO	320	236	15	8
AIFM3	YES	245	201	26	19
CRKL	YES	130	58	6	2
LZTR1	YES	363	331	31	61
P2RX6	NO	133	148	15	12
P2RX6P	NO	0	0	0	0
PI4KA	YES	742	461	83	23
SLC7A4	NO	244	279	9	7
SNAP29	NO	96	113	10	3
THAP7	YES	118	90	10	4
THAP7-AS1	NO	0	0	0	0
TMEM191A	NO	0	0	0	0
TUBA3FP	NO	0	0	0	0
HIC2	NO	284	192	9	0
MAPK1	YES	111	26	19	0
PI4KAP2	NO	0	0	0	0
PPIL2	NO	214	167	27	8
SDF2L1	NO	80	56	3	2
TMEM191C	NO	0	0	0	0
TOP3B	NO	331	213	27	7
UBE2L3	YES	46	3	5	0
YDJC	NO	123	74	4	4
YPEL1	NO	46	12	7	2
ZNF280B	NO	156	165	8	2
BCR	YES	561	408	38	2
FBXW4P1	NO	0	0	0	0
GNAZ	YES	182	50	7	0
RAB36	NO	154	119	16	10
RTDR1	NO	127	158	9	5
CABIN1	YES	812	703	68	25
CES5AP1	NO	0	0	0	0
CHCHD10	YES	47	33	3	5
DDT	NO	9	10	1	0
DDTL	NO	12	16	1	0
GGT5	NO	209	205	20	11
GSTT1	YES	88	49	7	5
GSTT2	NO	29	0	3	1
GSTT2B	NO	19	1	2	0
MIF	NO	46	25	2	2
SLC2A11	NO	191	185	16	17
SMARCB1	YES	159	43	17	0
ZDHHC8P1	NO	0	0	0	0
ZNF70	NO	161	169	10	3
ADORA2A	YES	177	121	8	2
ADORA2A-AS1	NO	0	0	0	0
BCRP3	NO	0	0	0	0
GGT1	NO	210	102	16	3
GUCD1	NO	84	77	8	4
SNRPD3	YES	54	9	7	0
SPECC1L	NO	387	304	41	6

Supplementary Table 5: Counts of missense and LGD deleterious coding variants seen in the 22q11.2 brain expressed genes retrieved from the denovo-db database.

GENE	22q11.2 Brain expressed 1st quartile	denovo-db_Missense_count	denovo.db_LGD_count
ARVCF	YES	0	2
C22orf29	NO	0	0
C22orf39	NO	0	0
CLDN5	YES	2	0
CLTCL1	NO	3	0
COMT	YES	1	0
DGCR10	YES	0	0
DGCR11	NO	0	0
DGCR14	NO	1	0
DGCR2	YES	2	0
DGCR5	YES	0	0
DGCR6	NO	0	0
DGCR8	YES	1	1
DGCR9	YES	0	0
GNB1L	NO	0	1
HIRA	NO	1	2
MRPL40	NO	0	0
PI4KAP1	NO	0	0
PRODH	YES	0	0
RANBP1	YES	1	0
SEPT5	YES	1	0
SLC25A1	YES	1	0
TANGO2	NO	0	0
TRMT2A	YES	2	0
TXNRD2	NO	0	0
UFD1L	NO	0	0
USP18	NO	0	0
ZDHHC8	YES	1	0
DGCR6L	YES	0	0
KLHL22	NO	0	0
MED15	YES	3	1
RTN4R	YES	2	0
SCARF2	NO	1	0
ZNF74	NO	0	0
AIFM3	YES	2	0
CRKL	YES	1	0
LZTR1	YES	8	2
P2RX6	NO	0	0
P2RX6P	NO	0	0
PI4KA	YES	1	0
SLC7A4	NO	2	0
SNAP29	NO	1	0
THAP7	YES	0	0
THAP7-AS1	NO	0	0
TMEM191A	NO	0	0
TUBA3FP	NO	0	0
HIC2	NO	0	0
MAPK1	YES	0	0
PI4KAP2	NO	0	1
PPIL2	NO	2	0
SDF2L1	NO	0	0
TMEM191C	NO	0	0
TOP3B	NO	3	0
UBE2L3	YES	0	0
YDJC	NO	0	0
YPEL1	NO	0	0
ZNF280B	NO	3	0
BCR	YES	2	0
FBXW4P1	NO	0	0
GNAZ	YES	1	0
RAB36	NO	0	0
RTDR1	NO	0	1
CABIN1	YES	4	0
CES5AP1	NO	0	0
CHCHD10	YES	2	0
DDT	NO	0	0
DDTL	NO	0	0
GGT5	NO	0	0
GSTT1	YES	0	0
GSTT2	NO	0	0
GSTT2B	NO	0	0
MIF	NO	0	1
SLC2A11	NO	0	0
SMARCB1	YES	0	1
ZDHHC8P1	NO	0	0
ZNF70	NO	0	0
ADORA2A	YES	0	0
ADORA2A-AS1	NO	0	0
BCRP3	NO	0	0
GGT1	NO	0	1
GUCD1	NO	0	0
SNRPD3	YES	0	0
SPECC1L	NO	2	1

Supplementary Table 6: Overall counts of missense and LGD deleterious coding variants seen in the 22q11.2 brain expressed genes retrieved from EXAc and denovo-db database.

	denovo-db		
	Missense	**LGD**	**Overall**
1st quartile	38	7	45
Other quartiles	19	8	27
	EXAc observed		
	Missense	**LGD**	**Overall**
1st quartile	4872	214	5086
Other quartiles	5038	318	5356
	EXAc expected		
	Missense	**LGD**	**Overall**
1st quartile	6706	556	7262
Other quartiles	6008	581	6589

Supplementary Table 7: Binomial test results obtained comparing EXAc observed variants distribution with expected values. Statistical analysis performed using GraphPad Software, Inc. CI of proportions were calculated using the GraphPad QuickCalcs Web site (accessed March 2018): http://www.graphpad.com/quickcalcs/ConfInterval1.cfm. In table legend as follow: #, count; % percentage; CI, confident intervals.

P value (one-tailed)	<0,001				
P value (two-tailed)	<0,001				
Outcome	Expected #	Observed #	Expected %	Observed %	95% CI of Observed %
22q11.2 first quartile brain expressed genes	7262	5086	69,55	48,71	47,75 to 49,67
Other 22q11.2 brain expressed genes	6589	5356	63,1	51,29	50,33 to 52,25
TOTAL	13851	10442	132,6	100	

Supplementary Table 8: denovo-db count of missense and/or LGD deleterious coding variants seen in genes harbored within 22q11.2 region.

Gene	Function Class	cDNA Variant	Protein Variant	Exon/Intron	CADD	rsID	ExAC Freq.	Main phenotype
LOC100996415	missense	c.617A>G	p.(D206G)	exon8	8.976	0	.0	autism
DGCR2	missense	c.190G>A	p.(A64T)	exon2	12.98	0	.0	congenital_heart_disease
DGCR2	missense	c.1105C>T	p.(H369Y)	exon8	27.0	rs200035453	.000725	developmentalDisorder
DGCR14	missense	c.479A>G	p.(N160S)	exon4	25.3	0	.0	autism
TSSK2	missense	c.707C>T	p.(P236L)	exon1	17.66	0	.000016	autism
GSC2	missense	c.410A>T	p.(Q137L)	exon2	24.4	0	.0	autism
SLC25A1	missense	c.874C>T	p.(R292W)	exon8	36.0	0	.0	schizophrenia
CLTCL1	missense	c.3923G>C	p.(G1308A)	exon25	15.771	0	.0	congenital_heart_disease
CLTCL1	missense	c.1941T>A	p.(N647K)	exon12	5.521	0	.0	developmentalDisorder
CLTCL1	missense	c.1921G>C	p.(V641L)	exon12	15.57	0	.0	schizophrenia
HIRA	stop-gained	c.2596C>T	p.(Q866*)	exon22	40.0	0	.0	developmentalDisorder
HIRA	missense	c.2639T>C	p.(M880T)	exon22	10.351	0	.0	autism
HIRA	frameshift	c.942_943del2	p.(C315Sfs*10)	exon10	-1.0	0	.0	congenital_heart_disease
CLDN5	missense	c.350T>G	p.(V117G)	exon1	23.9	rs199561265	.0	developmentalDisorder
CLDN5	missense	c.442C>A	p.(Q148K)	exon1	29.9	0	.0	developmentalDisorder
SEPT5	missense	c.358G>A	p.(E120K)	exon5	12.961	rs201694091	.000041	autism
GNB1L	frameshift	c.79_91del13	p.(H27Efs*41)	exon3	-1.0	0	.0	control
COMT	missense	c.250G>A	p.(E84K)	exon3	9.155	rs199690157	.000107	developmentalDisorder
ARVCF	stop-gained	c.1615C>T	p.(R539*)	exon8	42.0	0	.000008	autism
ARVCF	frameshift	NA	L809	14	-1.0	0	.0	intellectualDisability
DGCR8	frameshift	c.673del1	p.(V225Ffs*11)	exon2	-1.0	0	.0	autism
DGCR8	missense	c.211G>A	p.(G71R)	exon2	10.641	0	.000034	developmentalDisorder
TRMT2A	missense	c.226G>A	p.(E76K)	exon3	32.0	0	.000025	developmentalDisorder
TRMT2A	missense	c.1195G>T	p.(G399W)	exon7	32.0	0	.0	intellectualDisability
RANBP1	missense	c.227C>T	p.(S76F)	exon1	32.0	0	.000025	developmentalDisorder
ZDHHC8	missense-near-splice	c.557C>T	p.(T186I)	exon4	19.39	0	.0	control
SCARF2	missense	c.1152C>G	p.(C384W)	exon6	16.97	0	.0	developmentalDisorder
MED15	coding	c.612_644del33	p.(V204_Q215delinsV)	exon6	-1.0	0	.0	developmentalDisorder
MED15	missense	c.1618T>C	p.(S540P)	exon12	29.0	0	.0	developmentalDisorder
MED15	missense	c.2173G>A	p.(V725M)	exon17	21.0	0	.0	developmentalDisorder
MED15	missense	c.1697T>C	p.(M566T)	exon13	23.0	0	.0	intellectualDisability
PI4KA	missense	c.4939C>G	p.(P1647A)	exon41	15.74	0	.0	autism
SERPIND1	missense	c.118C>A	p.(P40T)	exon2	10.101	0	.0	developmentalDisorder
SERPIND1	missense	c.374G>A	p.(R125H)	exon2	35.0	0	.000016	developmentalDisorder
SERPIND1	frameshift	c.828_829insG	p.(A277Gfs*26)	exon2	-1.0	0	.0	intellectualDisability
SNAP29	missense	c.293T>C	p.(M98T)	exon2	18.851	0	.0	autism
CRKL	missense	c.832G>A	p.(E278K)	exon3	36.0	0	.0	control
AIFM3	missense	c.956C>T	p.(T319M)	exon11	23.9	rs139810844	.000016	autism
AIFM3	missense	c.1262A>G	p.(D421G)	exon14	25.601	0	.0	developmentalDisorder
LZTR1	frameshift	c.347_348insT	p.(R118Pfs*28)	exon4	-1.0	0	.0	control
LZTR1	missense	c.917C>T	p.(T306M)	exon9	17.941	0	.0	congenital_heart_disease
LZTR1	missense	c.742G>A	p.(G248R)	exon8	33.0	0	.0	congenital_heart_disease
LZTR1	missense	c.1988G>T	p.(G663V)	exon17	19.83	0	.0	autism
LZTR1	missense	c.2161G>A	p.(E721K)	exon18	14.37	0	.000016	control
LZTR1	missense	c.253G>A	p.(G85R)	exon2	27.701	0	.0	control
LZTR1	missense	c.64T>C	p.(S22P)	exon1	16.691	0	.0	bipolar_type2
LZTR1	missense	c.2074T>G	p.(F692V)	exon18	27.601	0	.0	developmentalDisorder
LZTR1	missense	c.731C>G	p.(S244C)	exon8	27.9	0	.0	developmentalDisorder
LZTR1	stop-gained	c.465C>G	p.(Y155*)	exon5	40.0	0	.0	developmentalDisorder
SLC7A4	missense	c.623T>C	p.(F208S)	exon2	21.3	0	.0	intellectualDisability
RIMBP3B	missense	c.4661G>T	p.(G1554V)	exon1	6.495	0	.0	autism
CCDC116	missense	c.1154G>C	p.(G385A)	exon4	14.95	0	.0	bipolar_type1
PPIL2	missense	c.960C>G	p.(I320M)	exon13	12.801	0	.0	congenital_heart_disease
PPIL2	missense	c.292G>A	p.(E98K)	exon6	19.83	0	.0	autism
TOP3B	missense	c.1996T>C	p.(C666R)	exon17	27.601	0	.0	autism
VPREB1	missense	c.190C>A	p.(P64T)	exon2	10.82	0	.0	intellectualDisability
ZNF280B	missense	c.893A>G	p.(Q298R)	exon4	14.421	0	.0	autism
ZNF280B	missense	c.1561A>C	p.(T521P)	exon4	9.426	0	.0	developmentalDisorder
ZNF280B	missense	c.535G>A	p.(E179K)	exon4	8.148	0	.000016	control
PRAME	missense	c.631A>G	p.(K211E)	exon5	13.86	0	.0	autism
RTN4R	missense	c.451C>T	p.(R151C)	exon2	15.461	0	.00001	autism
RTN4R	missense	c.157G>A	p.(V53M)	exon2	5.917	rs145292678	.00007	developmentalDisorder
IGLL5	missense	c.257G>A	p.(R86Q)	exon2	10.2	rs182236201	.000509	autism
GNAZ	missense	c.994A>G	p.(N332D)	exon3	24.8	0	.0	developmentalDisorder
BCR	missense	c.3203C>G	p.(P1068R)	exon19	14.601	0	.0	autism
BCR	missense	c.3223G>A	p.(V1075M)	exon19	12.11	0		autism
RTDR1	stop-gained	c.972C>G	p.(Y324*)	exon7	10.86	rs149525500	.000008	developmentalDisorder
RGL4	missense	c.1304C>G	p.(A435G)	exon10	13.561	0	.0	control
CHCHD10	missense-near-splice	c.43C>T	p.(R15C)	exon2	17.47	0	.0	autism
MMP11	missense	c.1415T>C	p.(L472P)	exon8	7.768	0	.0	developmentalDisorder
SMARCB1	codingComplex	c.1085_1087del3	p.(E362_K363delinsE)	exon8	-1.0	0	.0	developmentalDisorder
MIF	frameshift	c.28_29insT	p.(P11Afs*?)	exon1	-1.0	0	.0	schizophrenia
CABIN1	missense	c.5084G>T	p.(S1695I)	exon32	8.667	0	.0	congenital_heart_disease
CABIN1	missense	c.2903C>T	p.(A968V)	exon20	9.973	0	.0	schizophrenia
CABIN1	missense	c.5718G>T	p.(K1906N)	exon33	18.65	0	.0	autism
CABIN1	missense	c.1186C>T	p.(R396C)	exon10	25.701	0	.0	autism
SPECC1L	missense	c.392G>T	p.(R131L)	exon4	20.0	0	.000008	autism
SPECC1L	missense	c.2947A>G	p.(T983A)	exon12	16.361	rs143075516	.000016	autism
SPECC1L	codingComplex	NA	LDS345F	5	-1.0	0	.0	autism

Supplementary Table 9: 22q11.2 genes loci rearrangements' p-values retrieved from a CNV morbidity map analysis of 29,085 children with developmental delay/intellectual disability in comparison to 19,584 healthy controls. Statistical comparisons used the Peto and Peto modification of the Gehan-Wilcoxon test (because of non-proportional hazard ratios) to assess overall burden. Fisher's exact test was carried out to calculate the significance at specific thresholds (Suppl. Note, Coe BP et al., 2014). Statistical significative results are shown respectively highlighted in red for deletion and in green for duplication genomic events.

RefSeq Gene Name	Signature Dels (n=29085)	Control Dels (n=19584)	Signature Dels p-value	Signature Dups (n=29085)	Control Dups (n=19584)	Signature Dups p-value
USP18	6	1	0,15550904	88	15	1,3003E-08
GGT3P	6	5	0,74821535	74	71	0,98660163
DGCR6	6	73	1	74	331	1,00E+00
PRODH	282	79	8,8329E-14	134	350	1
DGCR5	286	81	1,0131E-13	138	358	1
DGCR9	199	72	1,63E-06	130	302	1
DGCR10	199	6	1,06E-36	130	40	3,1409E-06
DGCR2	165	4	9,62E-32	102	34	0,00013612
DGCR11	159	2	5,07E-33	99	17	1,82E-09
COMT	161	2	1,85E-33	100	17	1,27E-09
ARVCF	161	0	8,42E-37	100	18	3,42E-09
C22orf25	185	21	5,85E-22	100	13	1,44E-11
MIR185	161	0	8,42E-37	100	13	1,44E-11
DGCR8	175	2	1,56E-36	102	17	6,13E-10
MIR1306	161	0	8,42E-37	100	17	1,27E-09
TRMT2A	178	0	1,28E-40	106	18	3,97E-10
RANBP1	179	0	7,63E-41	107	22	1,18E-08
ZDHHC8	183	1	7,22E-40	108	26	2,07E-07
LOC150197	183	0	9,63E-42	106	30	5,42E-06
RTN4R	183	1	7,22E-40	106	31	9,87E-06
MIR1286	167	0	3,78E-38	102	29	8,78E-06
DGCR6L	147	1	7,07E-32	86	12	1,05E-09
PI4KAP1	147	2	2,15E-30	86	55	4,18E-01
RIMBP3	147	0	1,17E-33	86	16	7,03E-08
ZNF74	160	1	9,27E-35	99	16	6,5036E-10
SCARF2	159	1	1,54E-34	99	16	6,50E-10
KLHL22	159	1	1,54E-34	99	16	6,50E-10
MED15	159	1	1,54E-34	100	18	3,42E-09
POM121L4P	159	1	1,54E-34	98	16	9,41E-10
TMEM191A	159	1	1,54E-34	98	16	9,41E-10
PI4KA	166	2	1,48E-34	116	18	9,93E-12
SERPIND1	166	2	1,48E-34	116	16	1,02E-12
SNAP29	166	2	1,48E-34	116	16	1,02E-12
CRKL	166	2	1,48E-34	117	19	2,00E-11
AIFM3	166	2	1,48E-34	118	18	4,68E-12
LZTR1	166	2	1,48E-34	118	18	4,68E-12
THAP7	165	2	2,45E-34	118	18	4,68E-12
FLJ39582	165	2	2,45E-34	119	18	3,21E-12
MGC16703	165	2	2,45E-34	118	18	4,68E-12
P2RX6	165	2	2,45E-34	118	17	1,51E-12
SLC7A4	165	2	2,45E-34	118	17	1,51E-12
P2RX6P	165	2	2,45E-34	118	16	4,66E-13
LOC400891	165	3	5,59E-33	119	16	3,15E-13
POM121L8P	7	1	1,04E-01	21	219	1,00E+00
RIMBP3C	7	0	2,72E-02	21	383	1,00E+00
RIMBP3B	7	0	0,02721387	21	383	1
HIC2	19	1	0,00048733	26	388	1
PI4KAP2	15	0	0,0004422	20	390	1
RIMBP3B	15	0	0,0004422	20	383	0,99999667
RIMBP3C	15	0	0,0004422	20	383	0,99999667
UBE2L3	23	0	7,179E-06	24	6	0,01608546
YDJC	23	0	7,179E-06	18	0	9,4315E-05
CCDC116	23	0	7,18E-06	18	0	9,4315E-05
SDF2L1	23	0	7,18E-06	18	0	9,43E-05
MIR301B	23	0	7,18E-06	18	0	9,43E-05
MIR130B	23	0	7,18E-06	18	0	9,43E-05
PPIL2	23	0	7,18E-06	19	0	5,63E-05
YPEL1	23	1	7,37E-05	22	1	1,18E-04
MAPK1	23	1	7,37E-05	22	3	2,20E-03
PPM1F	24	1	4,57E-05	30	29	9,36E-01
TOP3B	37	27	6,74E-01	163	108	0,47445175
VPREB1	36	31	8,71E-01	20	4	0,01265948
LOC96610	36	29	0,80195056	20	4	0,01265948
ZNF280B	37	127	1	18	5	0,05146487
ZNF280A	37	122	1	18	5	0,05146487
PRAME	37	125	1	18	12	0,56788677
LOC648691	37	122	1	18	9	0,2997802
POM121L1P	25	117	1	11	8	0,65959636
GGTLC2	25	117	1	11	8	0,65959636
MIR650	29	195	1	25	52	0,99999964
IGLL5	27	101	1	23	17	0,67746915
RTDR1	20	3	0,00483221	23	3	0,00147201
GNAZ	20	2	0,00145037	23	23	0,93221785
RAB36	20	2	1,45E-03	24	3	0,00098181
BCR	20	3	4,83E-03	29	5	0,00125045
FBXW4P1	20	3	0,00483221	27	4	0,00094326
ZDHHC8P1	6	0	0,04554175	27	13	0,20228019
IGLL1	3	7	0,98719655	35	24	0,58302732
C22orf43	3	9	0,99704045	36	29	0,80195056
LOC91316	3	113	1	37	17	0,11928164
RGL4	3	0	0,21341842	30	11	0,05316876
ZNF70	3	0	0,21341842	31	11	0,04200979
VPREB3	3	0	0,21341842	31	10	0,02544494
C22orf15	3	0	0,21341842	31	10	0,02544494
CHCHD10	3	0	0,21341842	31	10	0,02544494
MMP11	3	0	0,21341842	32	10	0,01956709
SMARCB1	3	0	0,21341842	36	14	0,05024501
DERL3	3	0	0,21341842	35	15	0,08972506
SLC2A11	3	0	0,21341842	34	16	0,14789965
MIF	3	0	0,21341842	33	16	0,17454446
GSTT2B	3	82	1	29	133	1
GSTT2	3	82	1	29	133	1
DDTL	3	82	1	29	138	1
DDT	3	83	1	29	156	1
GSTT2	3	86	1	29	133	1
GSTTP1	3	311	1	29	320	1
LOC391322	3	705	1	29	296	1
GSTT1	3	706	1	29	296	1
GSTTP2	3	446	1	29	300	1
CABIN1	3	0	0,21341842	29	15	0,25078675
SUSD2	3	0	0,21341842	29	9	0,02497143
GGT5	3	0	0,21341842	30	9	0,01902381
POM121L9P	0	0	1	30	9	0,01902381
CYTSA	0	0	1	31	11	0,04200979
ADORA2A	0	0	1	31	11	0,04200979
C22orf45	0	0	1	31	11	0,04200979
UPB1	0	0	1	31	11	0,04200979
C22orf13	0	0	1	30	11	0,05316876
SNRPD3	0	0	1	30	10	0,03288304
GGT1	0	0	1	20	10	0,28229804
C22orf36	0	0	1	20	10	0,28229804
LOC644165	0	1	1	8	5	0,56733618
POM121L10P	0	1	1	8	3	0,29063187
PIWIL3	0	1	1	9	0	0,00971703
TOP1P2	0	0	1	9	0	0,00971703

Supplementary Table 10: Grouping list of significative 22q11.2 genes according to i), ii, and iii) criteria used for our enrichment analysis.

i) denovo-db	ii) CNV morbidity map	iii) Coe et al. model workflow	1st q. brain expressed
ADORA2A	ADORA2A	AIFM3	ARVCF
ADORA2A-AS1	AIFM3	ARVCF	CLDN5
AIFM3	ARVCF	CLDN5	COMT
ARVCF	BCR	DGCR8	DGCR10
BCR	C22orf15	LRRC75B	DGCR2
BCRP2	C22orf29	LZTR1	DGCR5
BMS1P20	C22orf39	MED15	DGCR8
CABIN1	CCDC116	RANBP1	DGCR9
CCDC116	CHCHD10	RTDR1	PRODH
CHCHD10	CLDN5	TRMT2A	RANBP1
CLDN5	CLTCL1	ZNF280B	SEPT5
CLTCL1	COMT		SLC25A1
COMT	CRKL		TRMT2A
CRKL	DGCR10		ZDHHC8
DDT	DGCR11		DGCR6L
DDTL	DGCR14		MED15
DGCR14	DGCR2		RTN4R
DGCR2	DGCR5		AIFM3
DGCR8	DGCR6L		CRKL
FAM230B	DGCR8		LZTR1
GGT1	DGCR9		PI4KA
GGT3P	FBXW4P1		THAP7
GGTLC2	GGT5		MAPK1
GNAZ	GNAZ		UBE2L3
GNB1L	GNB1L		BCR
GSC2	GP1BB		GNAZ
GSTT1	GSC2		CABIN1
GSTT2	HIC2		CHCHD10
GSTTP1	HIRA		GSTT1
GSTTP2	KLHL22		SMARCB1
HIRA	LZTR1		ADORA2A
IGLL1	MAPK1		SNRPD3
IGLL5	MED15		
LOC100996415	MIR1286		
LOC729444	MIR1306		
LRRC75B	MIR130B		
LZTR1	MIR185		
MED15	MIR301B		
MIF	MMP11		
MMP11	MRPL40		
PI4KA	P2RX6		
PI4KAP2	P2RX6P		
POM121L1P	PI4KA		
POM121L8P	PI4KAP1		
POM121L9P	PI4KAP2		
PPIL2	POM121L4P		
PRAME	POM121L9P		
RANBP1	PPIL2		
RGL4	PRODH		
RIMBP3B	RAB36		
RIMBP3C	RANBP1		
RTDR1	RIMBP3		
RTN4R	RIMBP3B		
SCARF2	RIMBP3C		
SEPT5	RTDR1		
SERPIND1	RTN4R		
SLC25A1	SCARF2		
SLC7A4	SDF2L1		
SMARCB1	SEPT5		
SNAP29	SERPIND1		
SNRPD3	SLC25A1		
SPECC1L	SLC7A4		
TANGO2	SMARCB1		
THAP7	SNAP29		
TOP3B	SNRPD3		
TRMT2A	SUSD2		
TSSK2	TBX1		
TXNRD2	THAP7		
USP18	TMEM191A		
VPREB1	TRMT2A		
YPEL1	TSSK2		
ZDHHC8	TXNRD2		
ZNF280B	UBE2L3		
	UFD1L		
	UPB1		
	USP18		
	VPREB1		
	VPREB3		
	YDJC		
	YPEL1		
	ZDHHC8		
	ZDHHC8P1		
	ZNF70		
	ZNF74		

Supplementary Table 11: List of 148 patients carrying a 22q11.2 distal deletion or duplication event, retrieved from 59 reports. For each proband are hereby reported the genomic rearrangement coordinates (in hg38), D- to G- genomic conventional interval defined between outer flanking LCRs22 (intervals extension defined in Suppl. Tab. 2), and inheritance status. Genomic rearrangement start- and stop-positions were uniformed to the Human Dec. 2013 (GRCh38/hg38) Assembly by using the hgLiftOver tool (http://www.genomaize.org/cgi-bin/hgLiftOver) or assuming the conventional LCRs22 boundary intervals whereas the event was diagnosed thought FISH analysis. In table abbreviations as follow: del, deletion; dup, duplication; NR, not reported; mat, maternal; pat, paternal.

Patient	LCR	Rearrangement	Position Start (hg38)	Position End (hg38)	Inheritance
Bassett_2017_1	F/G, G/H	dup	23333071	24663873	NR
Beddow_2011_1	D/E, E/F, F/G	del	21028664	24403449	de novo
Ben-Shachar_2008_1	D/E, E/F	del	21115711	24234031	de novo
Ben-Shachar_2008_2	D/E, E/F	del	21115711	24234031	de novo
Ben-Shachar_2008_3	D/E	del	21115711	22617529	de novo
Ben-Shachar_2008_4	D/E	del	21115711	22617529	NR
Ben-Shachar_2008_5	D/E	del	21115711	22617529	de novo
Ben-Shachar_2008_6	D/E	del	21115711	22617529	NR
Bosse_2014	D/E, E/F, F/G	del	21447315	24599033	NR
Bourdeaut_2011_1	E/F, F/G	del	23092773	23913813	NR
Bourdeaut_2011_2	E/F, F/G	del	23070483	24279892	NR
Bourdeaut_2011_3	D/E, E/F	del	21373831	24262652	NR
Bourdeaut_2011_4	D/E, E/F	del	21490991	24243092	NR
Bruce_2010_1	A/B, B/C, C/D, D/E	del	22061399	22485213	NR
Busse_2010_1	E/F	del	22617530	23307812	NR
Carvalho_2014_1	D/E	del	21603106	22256262O	NR
Chakrapani_2012_1	D/E, E/F, F/G	del	21444485	24234349	NR
Chang_2015_1	F/G, G/H	dup	23557777	24832122	pat
Coppinger_2009_1	F/G, G/H	dup	23307813	24684033	mat
Coppinger_2009_10	D/E, E/F, F/G	dup	21115711	24684033	mat
Coppinger_2009_2	F/G, G/H	dup	23307813	24684033	mat
Coppinger_2009_3	F/G, G/H	dup	23307813	24684033	mat
Coppinger_2009_4	F/G, G/H	dup	23307813	24684033	mat
Coppinger_2009_5	E/F, F/G, G/H	dup	23307813	24684033	mat
Coppinger_2009_6	E/F, F/G, G/H	dup	23307813	24684033	mat
Coppinger_2009_7	E/F, F/G, G/H	dup	23307813	24684033	NR
Coppinger_2009_8	D/E, E/F, F/G	dup	21115711	24304032	NR
Coppinger_2009_9	D/E, E/F, F/G	dup	21115711	23477813	de novo
D'Angelo_2014_1	E/F	del	22667530	23307813	mat
D'Angelo_2014_2	E/F	del	22717515	23357813	NR
D'Angelo_2018_1	D/E, E/F	del	21405283	23480738	NR
D'Angelo_2018_2	D/E	del	21114148	22617139	NR
D'Angelo_2018_3	E/F	del	22669599	23306640	MAT
D'Angelo_2018_4	E/F	del	22720693	23354277	NR
Demily_2017_1	F/G, G/H	dup	23397250	24592488	mat
Demily_2017_2	E/F, F/G, G/H	dup	222655814	2460105	de novo
Descartes_2008_1	E/F, F/G, G/H	dup	22617530	24684033	mat
Descartes_2008_2	E/F, F/G, G/H	dup	22617530	24684033	mat
Descartes_2008_3	E/F, F/G, G/H	dup	22617530	24684033	pat
Descartes_2008_4	E/F, F/G, G/H	dup	22617530	24684033	pat
Descartes_2008_5	E/F, F/G, G/H	dup	22617530	24684033	pat
Digilio_2015_1	D/E, E/F	del	21571972	23261302	de novo
Eaton_2011_1	D/E, E/F, F/G	del	21115711	24234031	mat
Fagerberg_2013_1	D/E, E/F, F/G	del	21405363	23319150	de novo
Fagerberg_2013_2	D/E, E/F	del	21562911	22620529	de novo
Fagerberg_2013_3	D/E, E/F	del	21562911	22620529	de novo
Hantash_2012_1	E/F, F/G, G/H	dup	23307813	24684033	de novo
Hantash_2012_2	E/F	dup	22617530	23307812	de novo
Jackson_2007_1	D/E, E/F, F/G	del	21115711	24234031	NR
Jackson_2007_2	D/E, E/F, F/G	del	21115711	24234031	NR
Jackson_2007_3	E/F, F/G, G/H	del	22617530	24684033	NR
Jackson_2007_4	F/G, G/H	del	23307813	24684033	MAT
Jackson_2007_5	F/G, G/H	del	23307813	24684033	NR
Kaufman_2016_1	D/E	del	21956976	22224590	NR
Lafay-Cousin_2009_1	D/E, E/F, F/G	del	21445284	24154605	NR
Lindgren_2015_1	D/E	del	21444416	22620448	NR
Lindgren_2015_2	D/E	del	21444416	22620448	NR
Lindgren_2015_3	D/E	del	21444416	22620448	NR
Lindgren_2015_4	D/E	dup	21444416	21876132	NR
Lindgren_2015_5	F/G, G/H	dup	23549586	24595724	NR
Lindgren_2015_6	E/F, F/G, G/H	dup	22712662	24595724	NR
Madan_2010_1	D/E, E/F	del	21585633	23421719	NR
Mikhail_2007_1	D/E, E/F	del	21115711	24234031	NR
Mikhail_2014_1	D/E	del	21954498	22224566	de novo
Mikhail_2014_10	E/F	del	22617530	23307812	de novo
Mikhail_2014_11	E/F	del	22617530	23307812	NR
Mikhail_2014_12	E/F	del	22617530	23307812	de novo
Mikhail_2014_13	E/F	del	22617530	23307812	NR
Mikhail_2014_2	D/E	del	21954498	22224566	NR
Mikhail_2014_3	D/E	del	21954498	22224566	NR
Mikhail_2014_4	D/E	del	21954498	22224566	de novo
Mikhail_2014_5	D/E	del	21954498	22224566	NR
Mikhail_2014_6	D/E	del	21954498	22224566	de novo
Mikhail_2014_7	D/E, E/F	del	21115711	24234031	NR
Mikhail_2014_8	D/E, E/F	del	21115711	24234031	NR
Mikhail_2014_9	D/E, E/F	del	21115711	24234031	de novo
Molck_2013_1	D/E	del	21111372	22619726	NR
Molck_2013_2	D/E	del	21111372	22656140	de novo
Molck_2015_1	F/G, G/H	dup	23549586	24595724	mat
Mosca_2016_1	D/E	del	21358465	22629925	de novo
Newbern_2008_1	D/E	del	21115711	22617529	de novo
Newbern_2008_2	D/E	del	21115711	22617529	de novo
Newbern_2008_3	D/E	del	21115711	22617529	de novo
Nguyen_2017_1	F/G, G/H	dup	23310330	24670505	NR
Nik-Zainal_2011_1	E/F	del	22915830	23300813	NR
Ou_2008_1	D/E	dup	21447315	22574114	pat
Ou_2008_2	D/E, E/F	dup	21447315	23285152	NR
Piccione_2011_1	E/F	del	22655530	23284813	de novo
Pinchefsky_2017_1	F/G, G/H	dup	23408742	24595702	de novo
Rauch_1999_1	D/E, E/F	del	21115711	24234031	mat
Rauch_2005_1	D/E, E/F	del	21115711	24234031	de novo
Rauch_2005_2	E/F	del	22617530	23307812	NR
Ribeiro-Bicudo_2013_1	F/G, G/H	dup	23353914	24564081	mat
Rødningen_2008_1	D/E	del	21115711	22617529	de novo
Rødningen_2008_2	D/E	del	21115711	22617529	de novo
Rodriguez-Santiago_2017_1	E/F, F/G	dup	22665537	24601506	de novo
Rodriguez-Santiago_2017_2	F/G, G/H	dup	23323169	24601506	NR
Saitta_1999_1	D/E	del	21115711	22617529	de novo
Sedghi_2015_1	E/F	dup	22617530	23307812	mat
Sgardioli_2017_1	D/E	del	21115711	22617529	NR
Shaikh_2007_1	E/F	del	22617530	23307812	de novo
Shimojima_2010_1	E/F, F/G, G/H	dup	22707515	24599033	de novo
Spineli-Silva_2017_1	D/E	del	21110474	22620492	NR
Stoll_2013_1	D/E	del	21954498	22224566	pat/mat
Stoll_2013_10	D/E	del	21954498	22224566	NR
Stoll_2013_11	D/E	del	21954498	22224566	NR
Stoll_2013_12	D/E	del	21954498	22224566	NR
Stoll_2013_13	D/E	del	21954498	22224566	NR
Stoll_2013_14	D/E	del	21954498	22224566	NR
Stoll_2013_15	D/E	del	21954498	22224566	NR
Stoll_2013_16	D/E	del	21954498	22224566	NR
Stoll_2013_17	D/E	del	21954498	22224566	NR
Stoll_2013_18	D/E	del	21954498	22224566	NR
Stoll_2013_19	D/E	del	21954498	22224566	NR
Stoll_2013_2	D/E	del	21954498	22224566	pat/mat
Stoll_2013_3	D/E	del	21954498	22224566	pat/mat
Stoll_2013_4	D/E	del	21954498	22224566	pat/mat
Stoll_2013_5	D/E	del	21954498	22224566	NR
Stoll_2013_6	D/E	del	21954498	22224566	NR
Stoll_2013_7	D/E	del	21954498	22224566	NR
Stoll_2013_8	D/E	del	21954498	22224566	NR
Stoll_2013_9	D/E	del	21954498	22224566	NR
Tan_2011_1	D/E, E/F, F/G, G/H	dup	21450428	27040327	pat
Tan_2011_2	E/F, F/G, G/H	dup	22722096	24536999	pat
Tan_2011_3	D/E, E/F	del	21441367	22617949	de novo
Tan_2011_4	D/E	del	21097948	24243101	de novo
Tan_2011_5	D/E	del	21968733	22215432	mat
Torti_2013	D/E, E/F	del	21444485	23062416	NR
Toth_2011_1	D/E, E/F, F/G	del	21454461	24247296	NR
Van Campenhout_2012_1	D/E	dup	21115711	22617529	de novo
Verhoeven_2011_1	D/E	del	21740983	22479776	de novo
Wieser_2005_1	D/E, E/F, F/G	del	21115711	24234031	de novo
Wincent_2010_1	F/G, G/H	dup	23325828	24596103	mat
Wincent_2010_10	E/F, F/G, G/H	dup	22689771	24599415	pat
Wincent_2010_2	F/G, G/H	dup	23330997	24581318	mat
Wincent_2010_3	F/G, G/H	dup	23330997	24683958	pat
Wincent_2010_4	F/G, G/H	dup	23342137	24595985	mat
Wincent_2010_5	F/G, G/H	dup	23345105	24558746	NR
Wincent_2010_6	F/G, G/H	dup	23365833	24683958	pat
Wincent_2010_7	F/G, G/H	dup	23325828	24582103	de novo
Wincent_2010_8	E/F, F/G, G/H	dup	22650368	24593969	NR
Wincent_2010_9	E/F, F/G, G/H	dup	226558I4	24600997	NR
Xu_2008_1	D/E	del	21115711	22617529	de novo
Yu_2011_1	D/E, E/F, F/G	del	21155711	22747513	NR
Yu_2011_2	D/E, E/F, F/G, G/H	del	21145711	24264032	NR
Yu_2011_3	D/E, E/F, F/G, G/H	del	21245711	24384032	NR
Yu_2011_4	E/F, F/G	del	22657530	23317813	de novo
Shi 2018	F/G	del	23951103	23988163	NR

Supplementary Table 12: HPO phenotypic description of 148 patients carrying a 22q11.2 distal deletion or duplication event, according to the clinical information retrieved from 59 reports.

Patient	HPO features
Bassett_2017_1	HP:0000152, HP:0000707, HP:0001626, HP:0000924, HP:0000478, HP:0002715, HP:0025031, HP:0000119, HP:0040064, HP:0000818, HP:0000598, HP:0002086, HP:0001507, HP:0001574, HP:0003549, HP:0001939, HP:0001871, HP:0025142, HP:0003011, HP:0001608, HP:0001197, HP:0100753, HP:0000708
Beddow_2011_1	HP:0000023, HP:0002021, HP:0000750, HP:0012758, HP:0000707, HP:0007018, HP:0100034, HP:0000722, HP:0001644, HP:0030692, HP:0100006, HP:0001249, HP:0012759, HP:0000276, HP:0000324, HP:0100807, HP:0000160, HP:0012745, HP:0001836, HP:0004467, HP:0011039,
Ben-Shachar_2008_1	HP:0001622, HP:0008897, HP:0000319, HP:0000175, HP:0002553, HP:0000924
Ben-Shachar_2008_2	HP:0001622, HP:0008897, HP:0001660, HP:0000319, HP:0002553, HP:0000307, HP:0000307, HP:0000218, HP:0000430
Ben-Shachar_2008_3	HP:0001622, HP:0008897, HP:0001647, HP:0000319, HP:0000490, HP:0000430
Ben-Shachar_2008_4	HP:0001622, HP:0008897, HP:0000319, HP:0002553, HP:0000924
Ben-Shachar_2008_5	HP:0001622, HP:0008897, HP:0000490, HP:0000430
Ben-Shachar_2008_6	HP:0008897, HP:0000319, HP:0002553, HP:0000307, HP:0000924, HP:0000490, HP:0000430
Bosse_2014	HP:0000717, HP:0001250, HP:0030692, HP:0009725, HP:0100836
Bourdeaut_2011_1	HP:0100836, HP:0100242
Bourdeaut_2011_2	HP:0100836, HP:0009726
Bourdeaut_2011_3	HP:0004375
Bourdeaut_2011_4	HP:0004375
Bruce_2010_1	HP:0008897, HP:0001518,HP:0004482, HP:0000325, HP:0000347, HP:0008929, HP:0004209, HP:0011968, HP:0000369, HP:0001270, HP:0001374, HP:0000430, HP:0001629, HP:0002650, HP:0000275, HP:0000347
Busse_2010_1	HP:0011608, HP:0410031, HP:0000316, HP:0000396, HP:0004322, HP:0001328, HP:0012759
Carvalho_2014_1	HP:0001622, HP:0001511, HP:0008897, HP:0000252, HP:0000601, HP:0000582, HP:0012745, HP:0000430, HP:0000219, HP:0000307, HP:0000377, HP:0000377, HP:0011842, HP:0001627, HP:0000708, HP:0001263, HP:0001249, HP:0001382
Chakrapani_2012_1	HP:0000356, HP:0030024, HP:0008897, HP:0001482, HP:0002664, HP:0000324, HP:0000319, HP:0002714
Chang_2015_1	HP:0001562, HP:0001788, HP:0001250, HP:0001302, HP:0002521, HP:0000252, HP:0002056, HP:0000486, HP:0000316, HP:0010805, HP:0001212, HP:0012759, HP:0012758, HP:0000365 HP:0006829,HP:0012294, HP:0000407, HP:0000365
Coppinger_2009_1	HP:0001263, HP:0000750, HP:0000252, HP:0001999, HP:0001252
Coppinger_2009_10	HP:0001263, HP:0000708, HP:0001999
Coppinger_2009_2	HP:0001263, HP:0000750, HP:0001999, HP:0001252
Coppinger_2009_3	HP:0001263
Coppinger_2009_4	HP:0001263, HP:0000750, HP:0000256
Coppinger_2009_5	HP:0001263, HP:0000750, HP:0011842, HP:0000252, HP:0001999, HP:0001252, HP:0001250
Coppinger_2009_6	HP:0011842, HP:0001263, HP:0001627, HP:0001999
Coppinger_2009_7	HP:0001263, HP:0000750, HP:0007018, HP:0000256, HP:0001999
Coppinger_2009_8	HP:0001263, HP:0000750, HP:0011842, HP:0001627, HP:0000252, HP:0001999, HP:0001252
Coppinger_2009_9	HP:0001263, HP:0000750, HP:0011843, HP:0001999
D'Angelo_2014_1	HP:0001252, HP:0001263, HP:0001250, HP:0000248, HP:0000490, HP:0000505, HP:0001388
D'Angelo_2014_2	HP:0001263, HP:0001249, HP:0001252, HP:0001344, HP:0000708, HP:0002360, HP:0001999, HP:0000486, HP:0000870, HP:0000054
D'Angelo_2018_1	HP:0001513, HP:0100836
D'Angelo_2018_2	HP:0001513
D'Angelo_2018_3	HP:0001513
D'Angelo_2018_4	HP:0001513
Demily_2017_1	HP:0002463, HP:0000750, HP:0000708, HP:0000718, HP:0000752, HP:0000729, HP:0000708,
Demily_2017_2	HP:0000708, HP:0001263, HP:0001249, HP:0012759, HP:0001251, HP:0002313, HP:0001152, HP:0000666, HP:0001272
Descartes_2008_1	HP:0001263, HP:0000750, HP:0000256, HP:0001999, HP:0001252, HP:0001249
Descartes_2008_2	HP:0001263,HP:0000750, HP:0000256, HP:0001999, HP:0001252, HP:0001249
Descartes_2008_3	HP:0001263,HP:0000750, HP:0001999, HP:0001249
Descartes_2008_4	HP:0001263, HP:0000750, HP:0001249
Descartes_2008_5	HP:0001263, HP:0000750, HP:0001249
Digilio_2015_1	HP:0002342, HP:0001510, HP:0008947, HP:0000252, HP:0001537, HP:0001999, HP:0000089, HP:0001195, HP:0000490, HP:0000494, HP:0000347, HP:0000233, HP:0009765, HP:0030682, HP:0001627, HP:0011623, HP:0001647, HP:0005301
Eaton_2011_1	HP:0004376, HP:0030680, HP:0000598
Fagerberg_2013_1	HP:0001622, HP:0000252, HP:0011555, HP:0001669, HP:0012758, HP:0012759, HP:0000582, HP:0000347,
Fagerberg_2013_2	HP:0004322, HP:0001629, HP:0001719, HP:0031477, HP:0005301, HP:0012758, HP:0012759, HP:0000582,
Fagerberg_2013_3	HP:0001622, HP:0001518, HP:0012758, HP:0012759, HP:0000582, HP:0000598,
Hantash_2012_1	HP:0000717, HP:0000729, HP:0012758, HP:0001249, HP:0000729
Hantash_2012_2	HP:0012758, HP:0000750, HP:0000729, HP:0000729
Jackson_2007_1	HP:0000175, HP:0002020, HP:0000582, HP:0000286, HP:0000377, HP:0000324, HP:0001182, HP:0008610, HP:0001263, HP:0001629, HP:0100006, HP:0100836
Jackson_2007_2	HP:0000175, HP:0002020, HP:0000582, HP:0001631, HP:0001629, HP:0001088, HP:0000272, HP:0004209, HP:0100006, HP:0100836
Jackson_2007_3	HP:0100006, HP:0100836
Jackson_2007_4	HP:0100006, HP:0100836
Jackson_2007_5	HP:0100006, HP:0100836
Kaufman_2016_1	HP:0000729, HP:0007018, HP:0010865, HP:0012169, HP:0001328, HP:0012759, HP:0000272, HP:0000601, HP:0000455, HP:0000154, HP:0000219, HP:0000232,HP:0001611, HP:0001999, HP:0000303, HP:0000709, HP:0000708, HP:0007302
Lafay-Cousin_2009_1	HP:0011332, HP:0000384, HP:0001140, HP:0030692, HP:0000396,HP:0000319, HP:0000219, HP:0002714,HP:0100006
Lindgren_2015_1	HP:0001622, HP:0001510, HP:0001511, HP:0100867, HP:0001252, HP:0000486, HP:0001270, HP:0012758
Lindgren_2015_2	HP:0001511, HP:0000276, HP:0000347, HP:0000407, HP:0007018, HP:0007302, HP:0012758, HP:0000750, HP:0012168, HP:0006919
Lindgren_2015_3	HP:0001511, HP:0001622, HP:0000252, HP:0000276, HP:0000272, HP:0000319, HP:0000347, HP:0000219, HP:0000218, HP:0000185, HP:0010865, HP:0007018, HP:0000708, HP:0012758, HP:0001263, HP:0001270, HP:0004691, HP:0100716, HP:0002099
Lindgren_2015_4	HP:0009905, HP:0000329, HP:0007018, HP:0000365, HP:0000407, HP:0001250, HP:0000487, HP:0000545, HP:0000119, HP:0004482, HP:0012450, HP:0001263, HP:0000750, HP:0012758,
Lindgren_2015_5	HP:0003119, HP:0001513, HP:0002099, HP:0012166, HP:0001249, HP:0012759, HP:0000545
Lindgren_2015_6	HP:0007018, HP:0008760, HP:0001250, HP:0000709, HP:0000739
Madan_2010_1	HP:0001622, HP:0001647, HP:0002616, HP:0005111, HP:0011623, HP:0001629, HP:0011649, HP:0004322, HP:0008591, HP:0000405, HP:0000365, HP:0002650, HP:0003423, HP:0001249, HP:0012759, HP:0000613, HP:0000545, HP:0001430, HP:0001999, HP:0001822, HP:0001252, HP:0012817
Mikhail_2007_1	HP:0007018, HP:0001513, HP:0001999, HP:0001256, HP:0001622
Mikhail_2014_1	HP:0001622, HP:0001518, HP:0001510, HP:0004322, HP:0001510, HP:0001263, HP:0100602, HP:0001249, HP:0002370, HP:0000252, HP:000199, HP:0040083, HP:0002007, HP:0000275, HP:0000276, HP:0000307, HP:0000369, HP:0000460, HP:0000426,
Mikhail_2014_10	HP:0001622, HP:0001263, HP:0001252, HP:0004482, HP:0001999, HP:0009237, HP:0002020, HP:0000316, HP:0011220,
Mikhail_2014_11	HP:0001263, HP:0001999, HP:0000490,
Mikhail_2014_12	HP:0001263, HP:0001249, HP:0000729, HP:0001999, HP:0001382,
Mikhail_2014_13	HP:0001263, HP:0001249, HP:0000717, HP:0001250, HP:0001999,
Mikhail_2014_2	HP:0001622, HP:0001518, HP:0001510, HP:0004322, HP:0001263, HP:0001562, HP:0001249, HP:0001250, HP:0001999,

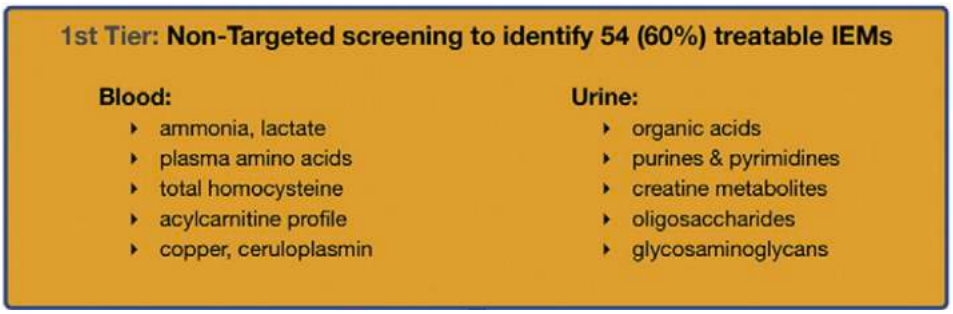
	HP:0001171, HP:0000871, HP:0001331, HP:0000824, HP:0010627, HP:0000609, HP:0000463, HP:0000325,
Mikhail_2014_3	HP:0001510, HP:0004322, HP:0001263, HP:0001249, HP:0001252, HP:0001250, HP:0001643, HP:0001999, HP:0004691, HP:0008734, HP:0011750, HP:0000054, HP:0011220, HP:0045025,
Mikhail_2014_4	HP:0001622, HP:0001510, HP:0001510, HP:0001263, HP:0001249, HP:0007018, HP:0000252, HP:0001999, HP:0000341, HP:0000664, HP:0012471, HP:0004428, HP:0000678,
Mikhail_2014_5	HP:0001622, HP:0001263, HP:0001249, HP:0007018, HP:0001250, HP:0001999,
Mikhail_2014_6	HP:0001622, HP:0001263, HP:0001249, HP:0001647, HP:0001999, HP:0004576,
Mikhail_2014_7	HP:0001622, HP:0001510, HP:0004322, HP:0001510, HP:0001263, HP:0006695, HP:0000252, HP:0001999, HP:0000664, HP:0002553, HP:0000582, HP:0000431, HP:0000385, HP:0000175,
Mikhail_2014_8	HP:0001622, HP:0001263, HP:0001249, HP:0007018, HP:0006695, HP:0001999, HP:0001513, HP:0000054, HP:0000823, HP:0000365, HP:0000402, HP:0002251, HP:0000618,
Mikhail_2014_9	HP:0001622, HP:0001518, HP:0001510, HP:0004322, HP:0001510, HP:0001263, HP:0001249, HP:0000752, HP:0001250, HP:0006695, HP:0001999, 01643, HP:0000252, HP:0000248, HP:0000568, HP:0000490, HP:0002744,
Molck_2013_1	HP:0001518, HP:0001513, HP:0040195, HP:0001263, HP:0012758, HP:0000315, HP:0000365, HP:0001999, HP:0004209,
Molck_2013_2	HP:0001518, HP:0008897, HP:0040195, HP:0011624, HP:0001655, HP:0000174, HP:0001263, HP:0012758, HP:0006695, HP:0001999
Molck_2015_1	HP:0000220, HP:0000369, HP:0000286, HP:0000430, HP:0002558, HP:0002650, HP:0001388, HP:0005180, HP:0001328, HP:0012759
Mosca_2016_1	HP:0000729, HP:0007018, HP:0000717
Newbern_2008_1	HP:0000347, HP:0001263, HP:0000252, HP:0004322, HP:0012020, HP:0011682,
Newbern_2008_2	HP:0000220, HP:0000347, HP:0001328, HP:0000252, HP:0004322, HP:0001660
Newbern_2008_3	HP:0000193, HP:0000347, HP:0001328, HP:0000252, HP:0004322, HP:0001660
Nguyen_2017_1	HP:0010775, HP:0004875, HP:0001518, HP:0000494, HP:0001195, HP:0001252, HP:0006610, HP:0009602, HP:0002032, HP:0002575, HP:0010049, HP:0001719, HP:0011652, HP:0008439,
Nik-Zainal_2011_1	HP:0040314, HP:0000013, HP:0008684, HP:0008724, HP:0010462, HP:0004736, HP:0025481, HP:0002948, HP:0002650, HP:0003468, HP:0003440, HP:0000202, HP:0000377, HP:0001631, HP:0005301, HP:0031297, HP:0009778, HP:0003422HP:0009660, HP:0009557, HP:0004207.
Ou_2008_1	HP:0001263, HP:0000325, HP:0002007, HP:0000337, HP:0000494, HP:0000486, HP:0100277, HP:0000411, HP:0000436, HP:0001212,
Ou_2008_2	HP:0002023, HP:0000081, HP:0000076, HP:0001655, HP:0031251, HP:0005556, HP:0000294, HP:0005280, HP:0001199, HP:0001800
Piccione_2011_1	HP:0001250, HP:0001249, HP:0001263, HP:0012758, HP:0012759 HP:0002123, HP:0002123
Pinchefsky_2017_1	HP:0001263, HP:0000729, HP:0002353, HP:0000219, HP:0000343, HP:0000414, HP:0002553, HP:0011474
Rauch_1999_1	HP:0011610, HP:0030680, HP:0011611, HP:0005374, HP:0008362, HP:0001800, HP:0000452, HP:0000356, HP:0012386
Rauch_2005_1	HP:0004502, HP:0000384, HP:0008053, HP:0012716, HP:0000486, HP:0012759, HP:0000316, HP:0000426, HP:0001629
Rauch_2005_2	HP:0001629, HP:0001642, HP:0011726, HP:0001643, HP:0003186, HP:0006610
Ribeiro-Bicudo_2013_1	HP:0001263, HP:0000750, HP:0000708, HP:0000202, HP:0001999, HP:0001249
Rødningen_2008_1	HP:0000319, HP:0000430, HP:0001622, HP:0001263, HP:0000407, HP:0000252, HP:0000119
Rødningen_2008_2	HP:0000319, HP:0000430, HP:0001622, HP:0001263, HP:0010296, HP:0000252, HP:0000377, HP:0011182, HP:0002353, HP:0000119
Rodriguez-Santiago_2017_1	HP:0000708, HP:0100753, HP:0001328, HP:0012759
Rodriguez-Santiago_2017_2	HP:0000708, HP:0100753, HP:0012759, HP:0008619, HP:0000407
Saitta_1999_1	HP:0000252, HP:0000316, HP:0030680, HP:0000193
Sedghi_2015_1	HP:0000185, HP:0001250, HP:0007281, HP:0000750, HP:0000341, HP:0000426, HP:0000286, HP:0000629, HP:0000490, HP:0000582, HP:0000534, HP:0000293, HP:0000347, HP:0001182, HP:0007018, HP:0002046, HP:0012759
Sgardioli_2017_1	HP:0030680, HP:0001611, HP:0000047, HP:0001629, HP:0000023 HP:0002650
Shaikh_2007_1	HP:0000252, HP:0000316, HP:0030680, HP:0000193
Shimojima_2010_1	HP:0001263, HP:0000750, HP:0007018, HP:0007018, HP:0001999, HP:0001250
Spineli-Silva_2017_1	HP:0000175, HP:0000384, HP:0001140, HP:0000272, HP:0100335, HP:0000154, HP:0000600, HP:0000347, HP:0000568, HP:0000324, HP:0000528, HP:0000327, HP:0000636, HP:0000152, HP:0000505, HP:0000528, HP:0000478, HP:0001249, HP:0002120, HP:0001274, HP:0002119, HP:0000717, HP:0002167, HP:0000707, HP:0008551, HP:0000370, HP:0000365, HP:0000413, HP:0008605, HP:0000359, HP:0000368, HP:0000598HP:0002650, HP:0000272, HP:0000772, HP:0003305, HP:0000347, HP:0008417, HP:0009601, HP:0002937, HP:0000327, HP:0000924, HP:0000505, HP:0000528, HP:0000478, HP:0001249, HP:0002120, HP:0001274, HP:0002119, HP:0000717, HP:0002575, HP:0025031, HP:0001601, HP:0002779, HP:0002089, HP:0002575, HP:0002086, HP:0000568, HP:0001140, HP:0001252, HP:0003011, HP:0000384, HP:0001574, HP:0009601, HP:0040064, HP:0004322, HP:0001507, HP:0000086, HP:0000104, HP:0000119, HP:0001636, HP:0001629, HP:0001626, HP:0008872, HP:0003011, HP:0000384, HP:0001574, HP:0009601, HP:0040064, HP:0004322, HP:0001507, HP:0000086, HP:0000104, HP:0000119, HP:0001636, HP:0001629, HP:0001626, HP:0008872,
Stoll_2013_1	HP:0100753, HP:0100543
Stoll_2013_10	HP:0100753
Stoll_2013_11	HP:0100753
Stoll_2013_12	HP:0100753
Stoll_2013_13	HP:0100753
Stoll_2013_14	HP:0100753
Stoll_2013_15	HP:0100753
Stoll_2013_16	HP:0100753
Stoll_2013_17	HP:0100753
Stoll_2013_18	HP:0100753
Stoll_2013_19	HP:0100753
Stoll_2013_2	HP:0100753, HP:0100543
Stoll_2013_3	HP:0100753, HP:0100543
Stoll_2013_4	HP:0100753, HP:0100543
Stoll_2013_5	HP:0100753
Stoll_2013_6	HP:0100753
Stoll_2013_7	HP:0100753
Stoll_2013_8	HP:0100753
Stoll_2013_9	HP:0100753
Tan_2011_1	HP:0001263, HP:0000750, HP:0000708, HP:0012443, HP:0000119, HP:0001627
Tan_2011_2	HP:0001263, HP:0000750, HP:0008610, HP:0001250
Tan_2011_3	HP:0004322, HP:0001518, HP:0000252, HP:0000582, HP:0011800, HP:0001622, HP:0000319, HP:0000219, HP:0000347, HP:0011332, HP:0001140, HP:0006695, HP:0001651, HP:0008947, HP:0003762
Tan_2011_4	HP:0004322, HP:0001518, HP:0000252, HP:0000582, HP:0011800, HP:0001622, HP:0000319, HP:0000219, HP:0000347, HP:0011968, HP:0000356, HP:0000708
Tan_2011_5	HP:0011968, HP:0000356, HP:0000708
Torti_2013	HP:0002006, HP:0000271, HP:0000324, HP:0003778, HP:0000154, HP:0004467, HP:0100335, HP:0000413, HP:0008513, HP:0000286, HP:0000545, HP:0012385, HP:0000774, HP:0012759
Toth_2011_1	HP:0100007, HP:0000384, HP:0002744, HP:0006695, HP:0001671
Van Campenhout_2012_1	HP:0001249, HP:0001270, HP:0001622, HP:0011968, HP:0001510, HP:0001263, HP:0007018, HP:0000717, HP:0001531
Verhoeven_2011_1	HP:0000739, HP:0000708, HP:0001629, HP:0001263, HP:0001643, HP:0000010, HP:0002205, HP:0002360, HP:0000722, HP:0025269, HP:0004322, HP:0001611, HP:0000490
Wieser_2005_1	HP:0100006, HP:0100836, HP:0001629, HP:0001651, HP:0000365, HP:0001263, HP:0012758

Wincent_2010_1	HP:0001263, HP:0000750, HP:0000708, HP:0001252
Wincent_2010_10	HP:0001263, HP:0012443, HP:0001999, HP:0001250
Wincent_2010_2	HP:0001263
Wincent_2010_3	HP:0001263
Wincent_2010_4	HP:0001263, HP:0001999
Wincent_2010_5	HP:0001263, HP:0000750, HP:0001999, HP:0001252
Wincent_2010_6	HP:0001263, HP:0000750
Wincent_2010_7	HP:0001263, HP:0000750
Wincent_2010_8	HP:0001263, HP:0000750, HP:0000708
Wincent_2010_9	HP:0001263, HP:0000750, HP:0012443, HP:0001999, HP:0001252, HP:0001250
Xu_2008_1	HP:0000175, HP:0000384, HP:0001140, HP:0000272, HP:0100335, HP:0000154, HP:0000600, HP:0000347, HP:0000568, HP:0000324, HP:0000528, HP:0000327, HP:0000636, HP:0000152, HP:0000505, HP:0000528, HP:0000478, HP:0001249, HP:0002120, HP:0001274, HP:0002119, HP:0000717, HP:0002167, HP:0000707, HP:0008551, HP:0000370, HP:0000365, HP:0000413, HP:0008605, HP:0000359, HP:0000368, HP:0000598HP:0002650, HP:0000272, HP:0000772, HP:0003305, HP:0000347, HP:0008417, HP:0009601, HP:0002937, HP:0000327, HP:0000924, HP:0000505, HP:0000528, HP:0000478, HP:0001249, HP:0002120, HP:0001274, HP:0002119, HP:0000717, HP:0002575, HP:0025031, HP:0001601, HP:0002779, HP:0002089, HP:0002575, HP:0002086, HP:0000568, HP:0001140, HP:0001252, HP:0003011, HP:0000384, HP:0001574, HP:0009601, HP:0040064, HP:0004322, HP:0001507, HP:0000086, HP:0000104, HP:0000119, HP:0001636, HP:0001629, HP:0001626, HP:0008872, HP:0003011, HP:0000384, HP:0001574, HP:0009601, HP:0040064, HP:0004322, HP:0001507, HP:0000086, HP:0000104, HP:0000119, HP:0001636, HP:0001629, HP:0001626, HP:0008872,
Yu_2011_1	HP:0000252, HP:0001999, HP:0012758, HP:0012759,
Yu_2011_2	HP:0000252, HP:0001999, HP:0012758, HP:0012759, HP:0001627, HP:0030060
Yu_2011_3	HP:0000252, HP:0001999, HP:0012758, HP:0012759, HP:0001627, HP:0000729, HP:0001249
Yu_2011_4	HP:0001250, HP:0001999, HP:0001263, HP:0012758, HP:0001531
Shi_2018	HP:0000220, HP:0001631, HP:0002788, HP:0001873

Supplementary Table 13: HPO nodes "0" overall features count matched to the corresponding genomic-driven LCRS22 and broken out *per* deletion and duplication event according to the clinical description of 148 patients retrieved carrying a 22q11.2 distal deletion or duplication event.

		22q11.2 distal duplication				22q11.2 distal deletion			
		D	E	F	G	D	E	F	G
Abnormality of head or neck	**HP:0000152**	15	36	40	30	207	105	36	4
Abnormality of the nervous system	**HP:0000707**	22	117	128	79	191	102	52	14
Abnormality of the eye	**HP:0000478**	3	3	6	6	32	20	2	0
Abnormality of the ear	**HP:0000598**	4	1	9	9	43	21	7	0
Abnormality of the voice	**HP:0001608**	0	0	1	1	3	0	0	0
Abnormality of the cardiovascular system	**HP:0001626**	5	5	9	8	63	50	14	2
Abnormality of the respiratory system	**HP:0002086**	0	0	4	4	14	0	1	0
Abnormality of the breast	**HP:0000769**	0	0	2	2	0	2	0	0
Abnormality of the digestive system	**HP:0025031**	3	1	3	3	16	7	4	0
Abnormality of the genitourinary system	**HP:0000119**	4	3	2	2	16	12	2	0
Abnormality of the skeletal system	**HP:0000924**	8	13	20	17	80	43	11	2
Abnormality of limbs	**HP:0040064**	2	2	4	4	16	13	4	0
Growth abnormality	**HP:0001507**	2	0	3	3	52	22	2	0
Abnormality of the musculature	**HP:0003011**	1	6	14	13	12	9	1	0
Abnormality of the integument	**HP:0001574**	4	3	3	3	18	11	4	0
Abnormality of connective tissue	**HP:0003549**	0	0	1	1	5	4	2	0
Abnormality of blood and blood-forming tissues	**HP:0001871**	0	0	1	1	0	0	1	0
Abnormality of the immune system	**HP:0002715**	0	0	2	2	1	1	4	0
Abnormality of the endocrine system	**HP:0000818**	0	0	1	1	5	2	0	0
Abnormality of metabolism/homeostasis	**HP:0001939**	0	1	2	2	0	0	0	0
Constitutional symptom	**HP:0025142**	0	0	1	1	0	0	0	0
Abnormality of prenatal development or birth	**HP:0001197**	1	0	4	4	27	12	1	0
Neoplasm	**HP:0002664**	1	0	0	0	21	26	27	7

Supplementary Table 14: Non-Targeted screening for ID treatable Inborn Errors of Metabolism (IEMs), performed in our patients (table adopted from van Karnebeek CD et al. 2014).



Supplementary Table 15: Specific primers used for RT-qPCR assays. a Forward Prime, b Reverse primer. Length of the amplicon in bp.

Gene	Sequences (5'-3')	Length
HPRT1	TGACACTGGCAAAACAATGCA[a] GGTCCTTTTCACCAGCAAGCT[b]	94
18S	CTACCACATCCAAGGAAGCA TTTTTCGTCACTACCTCCCCG[b]	71
IGLL5	CAATGGACTGGGGTGTACTG[a] CTCCTCAGGGGTCTCACAAC[b]	158
RTDR1	CCCAAGAGATCATCAGCAAAGa TCAGGACAAGCACCACATTGb	162
GNAZ	ATCCCGTGCTCCTTGTCTG[a] TGGTGCTCTTGCCTGAGTTG[b]	183
BCR	GGCAGGCAGAGGAGAGAAG[a] ACTGGGTGCTGGTGTCATC[b]	144
RAB36	GTGGTGGTTGGCGATCTCTAC[a] GAATCCCAGCAATCTCAAAGC[b]	133

Supplementary Table 16: Clinical overlap among NS patients reported carrying an heterozygotic LZTR1 mutations (adapted from Yamamoto et al. 2015). In table: NR: not reported; NA: not applicable; AoCo: aorta coarctation; ASD: atrial septal defect; BW: birth weight; PVS: pulmonary valve stenosis; SDS: SD score; VSD: ventricular septal defect; Mat: maternal; LVH: left ventricular hypertrophy; MVI: mitral valve insufficiency; MVP: mitral valve prolapse.

Study	**Yamamoto et al. 2015**					**Ghedira et al. 2017**
Variant	c.742G>A; p.G248R	c.850C>T; p.R284C	c.859C>T; p.H287Y	c.356A>G; p.+119C	c.740C>A; p.S247N	c.347C>T; p.A116V
Inheritance	Mat	Mat	de novo	de novo	Mat	de novo
Sex	Female	Female	Female	Female	Male	Male
Gestational age	Term	Term	Preterm	Term	Term	Term
BW (gr)	2270	2750	2130	3930	4000	NR
Length at birth (cm)	45	NR	47	52	53	NR
Typical facial features	+	+	+	+	+	+
Height (last evaluation: cm/ SDS for WHO-standard)	131.5/−2.1 SD	146/-1.8 SD	172.6/3.9 SD	164/3.2 SD	183/3.9 SD	93/-3 SD
Short/webbed neck	+	-	-	-	+	+
Pectus deformity	+	+	-	-	+	+
Cardiac abnormality	PVS/ASD	PVS	PVS/ASD	LVH	MVI	LVH/ASD
Cryptorchidism	NA	NA	+	NA	-	+
Renal abnormality	-	-	-	-	-	-
Abnormal hemostasis	-	+	+	-	-	-
Ophthamological abnormality	+	NR	+	-	-	-
Ectodermal findings	-	-	-	-	+	+
Curly hair	-	-	-	-	+	-
Sparse eyebrows	-	-	-	-	-	+
Hyperkeratosis pilaris	-	-	-	-	-	-
Ulerythema ophriogenes	-	-	-	-	-	+
Tumours	-	-	-	-	-	-
Developmental delay	-	-	+	-	+	+
Learning disability	-	-	+	-	+	NA
Other findings	Lacrimal duct obstruction, short stature and MVP in carriyng mother and grandfather			Lymphedema, varicose veins		

Supplementary Table 17: Clinical overlap among NS patients reported in Johnston JJ et al., 2018 and Nakaguma et al. 2018 carrying two inherited LZTR1 mutations both in homozygous or in compound heterozygote status. In table: NR: not reported; NA: not applicable; AoCo: aorta coarctation; ASD: atrial septal defect; BW: birth weight; PVS: pulmonary valve stenosis; SDS: SD score; VSD: ventricular septal defect; Mat: maternal; LVH: left ventricular hypertrophy; MVI: mitral valve insufficiency; MVP: mitral valve prolapse.

LZTR1 Variants	c.628C > T p.R210* pat; c.2220- 17C>A mat				2178C > A p.Y726* pat; c.1943- 256C>T ma		- 256C>T pat; c.1943- 256C>		p.E563Q pat; c.1687G > C		A>G pat; c.2090G > A p.R		c.27delG p.Q10fs pat; c.1149 +1G>A mat
Family/Proband ID as reported in study	1 II-1	1 II-2	1 II-3	1 II-4	2 II-4	2 II-5	3 II-1	3 II-4	4 II-1	4 II-2	5 II-4	5 II-1/2*	6 II-6
Sex	F	F	F	M	M	M	F	M	M	M	F	F	F
Prenatal hydrops, transl, or cardiac findings	–	+	–	+	+	+	+	+		+	+	+	–
BW (Kg/°C)	4.1 (90th)	3.9 (75-90th)	4.5 (>97th)	4.2 (90-97th)	3.3 (>97th)	4.5 (>97th)	3.32 (50-75th)	2.3 (AGA)	4.1	3.9 (90th)	3.46 (50th)	NR	90th
BL (cm/°C)	55.9 (>97th)	50.8 (50th)	52.7 (75-90th)	52 (50-75th)	43 (10-25th)	52 (95th)	50 (50-75th)			52 (70th)	48 (16th)		41 (50th)
BOFC (cm/°C)	U	35.6 (75-90th)	U	36.5 (90th)	34.5 (97th)	37(98th)	36 (>90th)		36	34.5 (50th)	36.2 (50-98th)		31.5 (90th)
Weight (Kg, last or onset evaluation)	22 (25th)	17.3 (10-25th)	16.1 (75th-90th)	7.6 (<3rd)	23-5 (30th)	12.5 (11th)	6.0 (<3rd)	12 (<3rd)	12 (10th)		13.5 (25th)		10.4 (<5th)
Ht (cm; last or onset evaluation)	121 (25th)	106 (10-25th)	98 (75-90th)	67 (<3rd)	113 (<3rd)	89 (3rd)	64 (<3rd)	85 (5-10th)	93 (25th)		89 (3rd)		78.2 (<5th)
OFC (cm; last or onset evaluation)	52 (50-75th)	52 (50-75th)	49 (75-90th)	45.5 (50th)	53 (57th)	50 (42nd)	44 (3rd)	49.5 (53rd)	51 (50-75th)		48.5 (25-50th)		
Ptosis	+	–	–	–	+	–		+	+		+		
Short nose or anteverted nares	–	+	-	-	-	-					-		
Depressed or wide bridge	–	-	-	+	-	+	+		+		+		+
Widely spaced eyes	–	-	-	-	+	-			+		-		+
Down-slanted palpebral fissures	–	-	-	-	+	-		+	+		+		+
Low-setears	–	+	-	+	+	-	+	+	+		+		
Post angulated ears	+	+	+	+	+	+			+		+		
Malformed ears	–	-	+	+	+	+			+		+		
Midface retrusion	+	+	-	-	+	+			-		+/ –		+
Micrognathia	+/ –	+/ –	+/ –	+/ –	+	+			+		+		
Broad/short neck	+	+	+	+	+	+	+		+	+	-	+	+
Low posterior hairline	+	+	+	+	+	-	+		+	-	-		
Wide-spaced nipples/broad chest	–	-	+	-	+	-	+				+	+	+
Pectus carinatum or excavatum	+	+	+	-	+	-			+		-		
Curly hair	-	+	+	-	-	-			-		+		
Cardiomyopathy	-	-	-	-	+	+	+	+	+	+	+		
CHD or valvular disease	+	+	-	+	+	+	-	+	+	+	+	+	
Cryptorchidism			NA	-	+	-			-	+			
Developmental delay/ID	-	+/ –	-	-	+	+			+		-		

Supplementary Table 17 (continuation): Clinical overlap among NS patients reported in Johnston JJ et al., 2018 and Nakaguma et al. 2018 carrying two inherited LZTR1 mutations both in homozygous or in compound heterozygote status. In table: NR: not reported; NA: not applicable; AoCo: aorta coarctation; ASD: atrial septal defect; BW: birth weight; PVS: pulmonary valve stenosis; SDS: SD score; VSD: ventricular septal defect; Mat: maternal; LVH: left ventricular hypertrophy; MVI: mitral valve insufficiency; MVP: mitral valve prolapse.

LZTR1 Variants	c.361C > G p.H121D pat; c.2264G > A p.R755Q mat	c.508C > T; c.614T > C p.R170W, I205T pat; c.508C > T; c.614T > C p.R170W; I205T mat	c.650A > C, p.E217A pat; c.650A > C, p.E217A mat	c.2062C > G, p.R688G pat; c.2062C > G, p.R688G mat	c.2325 +1G>A pat; c.1943- 256C>T mat	p.I821T pat; c.2462T > C		c.2212C>T:p.Gln738* pat; c.881G>T:p.Arg294Leu mat
Family/Proband ID as reported in study	7 II-1	8 II-11	9 II-2	10 II-3	11 II-6	12 V-1	12 V-4	Proband 1
Sex	M	M	F	F	M	F	F	M
Prenatal hydrops, transl, or cardiac findings	+			+	+	–	–	
BW (Kg/°C)	3.5 (AGA)	3.4 (AGA)	4 (AGA)	3.74 (75-90th)	2.6 (60th)	2.55 (AGA)	3.78 (AGA)	3.33 (60th)
BL (cm/°C)	50 (AGA)			51 (AGA)				50 (63th)
BOFC (cm/°C)	34 (AGA)			34 (25-50th)	31 (90th)			
Weight (Kg, last or onset evaluation)	21.4 (1-25th)	13.8 (10-25th)	13 Kg (50th)	14 (3rd)	3.1 (25th)	(25-50th)	12.7 (2nd)	
Ht (cm; last or onset evaluation)	113(<3rd)	88 (<3rd)	86 (3rd)	96.5 (<3rd)		(10-25th)	93.7 (<3rd)	123.6 cm (SDS –3.5)
OFC (cm; last or onset evaluation)	52 (25-50th)	51 (50-75th)	51 (97th)	48 (3rd-10th)		(50th)	47 (<3rd)	
Ptosis	–	+		–		+	+/–	+
Short nose or anteverted nares	-	-	+	-	+	+	-	
Depressed or wide bridge	-	-	+	+	+	+	-	
Widely spaced eyes	+	+	+	-		+	-	
Down-slanted palpebral fissures	+	+	+	-	+	+	-	
Low-setears	+	+		+	+	+	+	+
Post angulated ears	+	+		+	+	+	+	
Malformed ears	-	-		-		-	-	
Midface retrusion	+/–	-		-		+	+	
Micrognathia	-	-		+/–		-	-	+
Broad/short neck	+	+	+	+	+	-	-	
Low posterior hairline	+			-		+	+	
Wide-spaced nipples/broad chest	+			+	+	+	+	
Pectus carinatum or excavatum	+	-		-	+	+	+	+
Curly hair	+	+	-	-		+	-	
Cardiomyopathy	+	+	+	+	+	+	+	
CHD or valvular disease	+	+	+	-	+	+	+	+
Cryptorchidism	-	+						
Developmental delay/ID	+	+	+	+		+	+	