# Characterization of Intellectual disability and Autism comorbidity through gene panel sequencing

*Maria Cristina Aspromonte [1, 2], Mariagrazia Bellini [1, 2], Alessandra Gasparini [3], Marco Carraro [3], Elisa Bettella [1, 2], Roberta Polli [1, 2], Federica Cesca [1, 2], Stefania Bigoni [4], Stefania Boni [5], Ombretta Carlet [6], Susanna Negrin [6], Isabella Mammi [7], Donatella Milani [8], Angela Peron [9, 10], Stefano Sartori [11], Irene Toldo [11], Fiorenza Soli [12], Licia Turolla [13], Franco Stanzial [14], Francesco Benedicenti [14], Cristina Marino-Buslje [15], Silvio C.E. Tosatto [3, 16], Alessandra Murgia [1, 2], Emanuela Leonardi [1, 2]*

1. Molecular Genetics of Neurodevelopment, Dept. of Woman and Child Health, University of Padova, Padova, Italy
2. Fondazione Istituto di Ricerca Pediatrica (IRP), Città della Speranza, Padova, Italy
3. Dept. of Biomedical Sciences, University of Padova, Padova, Italy
4. Medical Genetics Unit, Ospedale Universitario S. Anna, Ferrara, Italy
5. Medical Genetics Unit, S. Martino Hospital, Belluno, Italy
6. Child Neuropsychiatry Unit, IRCCS Eugenio Medea, Conegliano, Italy
7. Medical Genetics Unit, Dolo General Hospital, Venezia, Italy
8. Pediatric Highly Intensive Care Unit, Department of Pathophysiology and Transplantation, University of Milano, Fondazione IRCCS, Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy
9. Child Neuropsychiatry Unit, Epilepsy Center, Santi Paolo-Carlo Hospital, Dept. of Health Sciences, University of Milano, Milano, Italy
10. Department of Pediatrics, Division of Medical Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA
11. Paediatric Neurology Unit, Dept. of Woman and Child Health, University Hospital of Padova, Padova, Italy.
12. Medical Genetics Department, APSS Trento, Trento, Italy
13. Medical Genetics Unit, Local Health Authority (ULSS2), Treviso, Italy
14. Genetic Counseling Service, Dept. of Pediatrics, Regional Hospital of Bolzano, Bolzano, Italy
15. Bioinformatics Unit, Fundación Instituto Leloir, Av. Patricias Argentinas 435, Buenos Aires, Argentina
16. CNR Institute of Neuroscience, Padova, Italy

**ABSTRACT**

Intellectual disability (ID) and autism spectrum disorder (ASD) are clinically and genetically heterogeneous diseases. Recent whole exome sequencing studies indicated that genes associated with different neurological diseases are shared across disorders and converge on common functional pathways. Using the Ion Torrent platform, we developed a low-cost next generation sequencing (NGS) gene panel that has been transferred into clinical practice, replacing single disease gene analyses for the early diagnosis of individuals with ID/ASD. The gene panel was designed using an innovative *in silico* approach based on disease networks and mining data from public resources to score disease-gene associations. We analyzed 150 unrelated individuals with

1

ID and/or ASD and a confident diagnosis has been reached in 26 cases (17%). Likely pathogenic mutations have been identified in another 15 patients, reaching a total diagnostic yield of 27%. Our data also support the pathogenic role of genes recently proposed to be involved in ASD. Although many of the identified variants need further investigation to be considered disease-causing, our results indicate the efficiency of the targeted gene panel on the identification of novel and rare variants in patients with ID and ASD.

**Keywords**

Gene panel, NGS, ID, ASD, variant interpretation, comorbidity

**INTRODUCTION**

Neuro-developmental disorders (NDDs) are common conditions including clinically diverse and genetically heterogeneous diseases. Intellectual disability (ID) is the most common NDD disorder, with a prevalence varying between 0.5 to 3% in general population, depending on patient and parent age, or the measure of intellectual quotient used (Leonard et al., 2011). ID is characterized by deficits in both intellectual and adaptive functioning that first manifest during early childhood. Children with intellectual disability (ID) exhibit increased risk to present potential co-occurring developmental conditions, such as autism spectrum disorders (ASDs) (28%), epilepsy (22.2%), stereotypic movement disorders (25%), and motor disorders, which substantially affect daily living and well-being (Almuhtaseb et al., 2014; Jensen and Girirajan, 2017; Kazeminasab et al., 2018). ASD in particular is characterized by deficits in social communication and interactions, as well as by repetitive behaviors and restrictive interests, is associated with poorer psychosocial and family related outcomes than ID alone (Totsika et al., 2011). ASD as well as epilepsy commonly coexist in specific neurodevelopmental disorders with ID, such as Fragile-X and Rett syndromes, or in phenotypes associated with specific copy number variations (CNVs) and single gene mutations. These make the differential diagnosis among these disorders extremely difficult based only on clinical features. Furthermore, it seems that patients affected by one of these disorders have high risk to develop other comorbid NDDs. Exome sequencing studies of family trios with ID, ASD, and epilepsy have revealed a significant excess of *de novo* mutations in probands, when compared to the normal population, and yielded a rich source of candidate genes contributing to these neurodevelopmental defects (Neale et al., 2012; Epi4K Consortium et al., 2013; Fromer et al., 2014). It has been estimated that mutations in more than 1,000 different genes might cause ID (Chiurazzi and Pirozzi, 2016). Both common and rare genetic variants in up to 1,000 genes have been linked to increased ASD risk (SFARI database, https://gene.sfari.org/). However, significant numbers of genes harboring *de novo*

mutations are shared across different neurodevelopmental or neuropsychiatric disorders (Vissers et al., 2010; Cukier et al., 2014). Many genes have already been shown to cause both ID and/or ASD, including *PTCHD1*, *SHANK3, NLGN4, NRXN1, CNTNAP2, UBE3A, FMR1, MECP2*, and others (Harripaul et al., 2017). Despite the apparent distinct pathogenesis for these disorders, analysis of network connectivity of the candidate genes revealed that many rare genetic alterations converge on a few key biological pathways (Vissers et al., 2010; Krumm et al., 2014). In the case of ASD, diverse integrative systems biology approaches highlighted how disease genes cluster together in networks enriched in synaptic function, neuronal signaling, channel activity, and chromatin remodeling (Gilman et al., 2012; O'Roak et al., 2012; Pinto et al., 2014). Accordingly, many ASD genes are synaptic proteins, chromatin remodelers, or FMRP targets, i.e. genes encoding transcripts that bind to FMRP (Iossifov et al., 2012). Alterations of the same classes of protein functions and biological processes involved in neuronal development, such as the mammalian target of rapamycin (mTOR) pathways, GABA receptor function or glutamate NMDA receptor function, have been also found implicated in intellectual disability, epilepsy, and schizophrenia (Endele et al., 2010; Gilman et al., 2011; Paoletti et al., 2013; Cristino et al., 2014; Krumm et al., 2014; Reijnders et al., 2017). The multiple genes and molecular pathways shared by ID, ASD and other developmental or psychiatric disorders indicate a common origin that explains the co-occurrence of these conditions (Barabási et al., 2011; Cukier et al., 2014).

Based on the hypothesis that common functional pathways explain comorbidity between diverse NDDs disorders, we developed an efficient and cost-effective amplicon-based multigene panel to assess the pathogenic role of genes involved in ID and ASDs comorbidity. The 74-gene panel was designed using an innovative *in silico* approach based on disease networks and mining data from public resources to score disease-gene association. Here, we present the genetic findings after applying this panel to 150 individuals from our cohort of individuals with ID and/or ASD, most of them were negative for array-CGH (aCGH), Fragile-X test and other specific genetic analyses (*MECP2*, *CDKL5*, *UBE3A*, chr15q methylation test, etc.). We adopted a manual prioritization procedure based on expert knowledge related to the disease phenotype and gene functions, which allowed detecting a causative or likely pathogenic variant in 27% of these patients. We describe diagnosed cases that highlight the critical steps of variant interpretation, in the clinical diagnostic context of neurodevelopmental conditions such as ID and ASDs. For each tested individual, we report a clinical description and genetic data from the 74 genes providing a set of genotype − phenotype associations, which can be used to train or test computational methods for prioritization of potential disease-causing variants.

## MATERIALS AND METHODS

### Patient selection

Patients were referred from clinical geneticists of 17 Italian public hospitals with a diagnosis of non-specific neurodevelopmental disorder. Clinical data were collected with a standardized clinical record describing clinical and family history, clinical phenotype (auxological parameters, neurological development, physical features, and behavioral profile), and presence of associated disorders. Data from neurophysiological profiles, electroencephalograms (EEG) and brain magnetic resonance imaging (MRI) were also collected. Table 1 summarizes the clinical data of the patients. Written informed consent was obtained from the patient's parents or legal representative. This study was approved by the Local Ethics Committee, University-Hospital of Padova, Italy.

### Gene panel selection

For the construction of an efficient and low-cost gene panel, we selected the most promising ID and/or ASD genes gathering data from public databases (AutismKB http://autismkb.cbi.pku.edu.cn, and SFARI https://sfari.org/resources/sfari-gene, Abrahams et al. 2013), OMIM, and PubMed. Candidate genes were extracted in particular from recent exome sequencing and meta-analysis studies (Supplementary Table S1). We collected a list of 972 genes scored according to recurrence in different sources, annotated for clinical phenotype, gene function, subcellular localization, and interaction with other known causative genes. Separated lists were generated considering ASD or ID association. The extracted information was stored in a dedicated SQL database used in conjunction with the disease network construction. Using data from STRING 9.0 (Franceschini et al., 2013), a disease protein-protein interaction (PPI) network was built starting from 66 high confidence genes (intersection list), shared by both ASD and ID gene lists. Emerging features of the network were assessed by enrichment analysis with Enrich web-server (Kuleshov et al., 2016). The same list was used as training set for Endeavour gene prioritization (https://endeavour.esat.kuleuven.be/) (Tranchevent et al., 2016). Hub direct interactors (i.e. genes with STRING degree score above 0.45) belonging to the top ranking prioritized list, but not included in the intersection, were also included in the most promising candidate gene list. To this list, we also added the top ranked genes associated to ID or ASD only (i.e. genes with at least 5 evidences for ID or ASD) (Supplementary Figure S1). The final panel set resulted in a manually curated list of 74 genes, comprising selected known causative genes, top-ranked genes by gene prioritization, and genes meeting PPI network parameters (Supplementary Table S2).

**Gene Panel Sequencing**

Nucleic acids were extracted from blood samples using the Wizard genomic DNA Promega Kit (Promega Corporation). Multiplex, PCR-based primer panels were designed with Ion AmpliSeq™ Designer (Thermo Fisher Scientific) to amplify all exons and flanking regions (10 bp) of the 74 selected genes (Thermo Fisher Scientific). Template preparation and enrichment were performed with the Ion One Touch 2 and Ion One Touch ES System, respectively (Thermo Fisher Scientific). Read alignment to the human genome reference (hg19/GRCh37) and variant calling were performed with the Ion Torrent Suite Software v5.02 (Thermo Fisher Scientific).

**Variant filtering**

An in house pipeline was built to create a database of genetic variants identified in our cohort and to annotate them with features provided by ANNOVAR, i.e. allelic frequency (AF) in control cohorts, variant interpretation from InterVar automated, ClinVar report, pathogenicity predictions and conservation scores. Detected variants were ranked for their frequency in the gnomAD (Lek et al., 2016), 1000G (1000 Genomes Project Consortium et al., 2015), and ExAC (Kobayashi et al., 2017) databases, as well as in our in house database of 150 patients. We excluded SNVs found more than twice in our cohort or reported with an AF higher than expected for the disorder, which has been calculated to be < 0.002% and <0.45% for autosomal dominant and recessive genes, respectively (Piton et al., 2013). Variants reported as risk factors for autism in the literature were nevertheless considered for further segregation analysis even if their frequency in the control populations exceeded the ID incidence. Rare variants were ranked for their pathogenicity prediction considering the consensus among twelve computational methods. ANNOVAR provides predictions from SIFT (Sim et al., 2012), the Polyphen-2 (Adzhubei et al., 2013) HDIV and HVAR versions, LRT (Chun and Fay, 2009), Mutation Taster (Schwarz et al., 2014), MutationAssessor (Reva et al., 2011), FATHMM (Shihab et al., 2014); PROVEAN (Choi and Chan, 2015), MetaSVM (Dong et al., 2015), MetaLR (Dong et al., 2015), M-CAP (Jagadeesh et al., 2016), fathmmMKL (Shihab et al., 2013, 2014) and CADD (Kircher et al., 2014). Conservation was evaluated with the scoring schemes GERP++ (Davydov et al., 2010), PhyloP (Pollard et al., 2010) and SiPhy (Garber et al., 2009). Intronic or synonymous variants near the exon-intron junction were also evaluated *in silico* for their impact on splicing using Human Splicing Finder (Desmet et al., 2009). The Integrated Genome Viewer platform (Robinson et al., 2011) has been used to exclude sequencing or alignment errors around selected SNVs.

**Variant validation and functional assays**

Selected variants were validated by Sanger sequencing. Segregation analysis was performed in the patient relatives when DNA samples were available. For apparent *de novo* variants, paternity and maternity were confirmed by the inheritance of rare detected variants in parental samples. In other cases pedigree concordance was checked using polymorphic microsatellite markers of chr15q described in (Giardina et al., 2008). For maternally inherited X-linked variants, the X-inactivation pattern of the mother was evaluated on the highly polymorphic androgen receptor (ARA locus) at Xq11-q12, as described in (Bettella et al., 2013). The X-inactivation was classified as random (ratio < 40:60) or significantly skewed (ratio ≥ 80:20).

Analysis of the transcripts was performed to confirm putative splicing variants. RNA was extracted from patient peripheral blood leukocytes and real-time PCR performed using random primers. cDNA was used as template in nested PCR reactions with specific primers in order to amplify the regions containing the mutation. PCR products were tested on 1.5% agarose gel and sequenced.

**Variant classification**

A clinical interpretation of selected variants was first evaluated using the InterVar (Li and Wang, 2017), adjusting the automated classification with our own findings. The American College of Medical Genetics and Genomics (ACMG) criteria (Richards et al., 2015) used to classify the variants are reported for both causative and likely pathogenic variants (Tables 2 and 3). All the identified variants have been submitted to LOVD database.

To support the disease relevance of each selected variant, multiple lines of evidence were required: conservation, allele frequency in population databases, computational inferences of variant effect, mode of inheritance, X-inactivation pattern, and disease segregation.

***In silico* analysis of candidate variants**

Canonical protein sequences were retrieved from UniProt (Apweiler et al., 2004) and protein domains predicted by InterPro (Mitchell et al., 2018). To evaluate conservation, orthologous sequences were downloaded from OMA Browser (Schneider et al., 2007) and aligned with MAFFT (Katoh and Standley, 2013). When available, crystal structures were retrieved from PDB (Rose et al., 2017). Structures of protein domains were modelled with MODELLER (Alva et al., 2016) (automatic best template selection), using templates predicted by HHpred (Alva et al., 2016). Structure of the CASK L27 domain and its complex with SAP97 have been analyzed using Mistic2 (Colell et al., 2018) to evaluate covariation between residues. Structures were manually explored with Pymol (Janson et al., 2017) or UCSF Chimera (Pettersen et al., 2004).

Disorder content and the presence of short linear motifs for protein interactions were assessed combining MobiDB (Potenza et al., 2015) and ELM (Gibson et al., 2015), using the interactive exploration tool ProViz (Jehl et al., 2016).

## RESULTS

### Gene panel description

The computational approach adopted to select panel genes includes genes recurrently mutated in ID or ASD conditions, genes shared among ID and ASD disease networks, and genes directly connected to ID/ASD. Of the 74 selected genes, 42 are FMRP targets, 21 postsynaptic proteins, and 16 chromatin modifiers. The majority of the selected genes are associated with autosomal dominand diseases; of these *DEAF1*, *PTEN* and *RELN* are also associated with a recessive disorder. Moreover, the panel includes 8 genes associated with autosomal recessive disorders and 19 genes associated with X-linked diseases. Twenty genes are associated with non-specific ID, while 40 genes are responsible for defined genetic syndromes. Seven genes have been found to confer autism susceptibility (*SHANK2, CNTNAP2, RELN, CHD8, NLGN3, NLGN4X, and PTCHD1*). At the time of panel design, for 11 of the selected genes there was only scant evidence in the literature about their association with ID and/or ASD (*ASH1L, NTNG1, KATNAL2, MIB1, MTF1, MYH10, PTPN4, TANC2, TBR1, TRIO*, and *WAC*). Two more genes associated with an OMIM ID have been meanwhile reclassified and the disease association confirmation is pending (*HDAC4* and *KIRREL3*) (Supplementary Table S2).

### Data Output and Quality

Our strategy allowed generating 263 reads per amplicon with 97% of them on target. The mean target region coverage for all patients ranged between 67 and 612. About 94% of target regions were covered at least 20x and 93% of amplicons had no strand bias. The uniformity of amplicon coverage resulted being 89%. A small proportion of targeted regions was weakly covered (<20x) throughout all patients. These are mainly first exons or GC-rich regions representing a well-known burden in an amplicon based strategy. Specifically, exon 4 of *MECP2*, exon 5 of *ARX*, exon 1 of *FMR1*, part of exon 21 of *SHANK3* have a read depth <10x. Depending on phenotype manifestations, otherwise these regions have been covered by Sanger sequencing, as they could not be analyzed reliably.

### Cohort description and diagnostic yield

Our cohort of 150 individuals is enriched in males (61%). 81.3% were sporadic cases and the remaining had a family history of neurodevelopmental disorders with siblings affected in 3.3%

of cases. At the time of molecular testing, the age of the patients was ranging between 2 and 42, with a median of 11 years. Although the vast majority of patients have ID, for four individuals clinicians did not report information about the presence of an intellectual defect. In 38 cases, the level of cognitive impairment has not been evaluated. Among the patients with a cognitive impairment evaluation, a slightly higher proportion have a moderate form (25.3%) than severe (22%) and mild ones (24.6%). 93 patients (62%) had both ID and ASD. 53 (35.3%) had only ID, while 4 (2.6%) had ASD and no information was provided about the presence of ID. Epilepsy was reported for 55 (36.6%) individuals, of which 6 had early onset epilepsy (<24 months). 39 patients with ID and ASD present also epilepsy (26%) (Table 1). MRI and EEG abnormalities were reported, respectively, for 37.3% and 24.6% of the sequenced individuals (Table 1). In most of the patients a structural pathogenic alteration of chromosomes was excluded by FISH, karyotype, or aCGH; however, in 17 of these, aCGH analysis revealed a copy number variations (CNVs) inherited from unaffected parents or involving gene-poor regions (Supplementary Table S3). 86 patients had a negative Fragile-X test, and 57 resulted negative to other single gene tests, such as *MECP2*, *CDKL5*, *UBE3A*, and/or to the Chr15q methylation test.

**Variant detection and prioritization**

In coding exons or exon-intron boundaries regions of the 74 genes, we detected on average about 74 single nucleotide variants (SNVs), with a range of 60 to 80 SNVs per patient. Overall, 202 coding or splicing SNVs passing the quality control (GQ>30, DP>20) and frequency filters (MAF<1%), were observed only once in our cohort. Based on the prioritization criteria described in the Methods section, we selected about 170 variants for further analysis, 47 of which were absent from the general population (Supplementary Table S4). We detected certainly causative mutations in 26 patients, leading to an overall diagnostic yield of 17.3% for the entire cohort. These rare or novel variants were predicted pathogenic by several approaches, found to be *de novo* or inherited from affected parents and consistent with the expected disease for the respective gene (Table 2). In other 15 patients, we identified 17 rare variants (missense, synonymous or splicing) with a putative although not established pathogenic role (Table 3). For these variants, either there were no family members available for segregation analysis, or the clinical features did not fit those that would have been expected for the respective gene. With the availability of functional studies or further sequencing data, we expect that a number of these variants will turn out to be benign, but some variants might be proven pathogenic. In particular, five of these putative pathogenic variants meet most of the ACMG criteria (p.Tyr2418Phe in *ATRX*, p.Gly1465Arg in *CREBBP*, c.730+1G>A in *DEAF1*.Val821Leu in *GRIN2B*, p.Pro576Gln

in *SHANK2*) and thus with segregation analysis we would be able to assign their causative role on the proband's phenotype. For two other cases, MR1769.01 and MR2053.01, either the digenic or autosomal recessive transmission is possible, only with functional assays supporting the pathogenic effect of the identified variants. In addition to a novel missense mutation in *FOXP1* gene, MR1769.01 carries a rare variant in *CNTNAP2,* which is transcriptionally regulated by *FOXP1*. As previously proposed by O'Roak and collaborators, we hypothesize a two-hit model for the disease risk in this patient, where a mutant *FOXP1* protein leads to an amplification of the deleterious effects of p.Asp417Tyr in *CNTNAP2* (O'Roak et al., 2011). In MR2053.01, we detected two variants in GAD1 gene, which in associated with an autosomal recessive phenotype. The paternally inherited missense p.Leu284Phe maps on the pyridoxal 5'-phosphate (PLP) transferase domain and is predicted as pathogenic by several computational methods, while the maternally inherited c.83-3T>C is predicted to alter splicing mechanisms. However, we were not able to demonstrate its pathogenicity by qualitative analysis of the transcript extracted from the mother.

After segregation analysis, we were able to classify about one hundred of the selected variants as likely benign. The majority of these variants were found in genes associated with highly penetrant disorders inherited from apparently healthy parents (n=85) or found in individuals with a causative mutation in another gene consistent with the proband phenotype (n=7). Five variants in X-linked genes were inherited from a healthy father or the X-inactivation pattern did not supported their pathogenic role (Supplementary Table S4).

Finally, some rare or novel inherited variants with strong pathogenicity predictions were classified as possible contributing factors. These variants occur on genes known to confer autism susceptibility (e.g. *SHANK2*, *RELN*, *CNTNAP2*) or that have been reported in the literature in individuals with very mild phenotype (e.g. *TRIO*, and *SLC6A1*) (Supplementary Table S4).


**Disea**se **causing variants**

We identified likely causative variants in 26 patients (17%) and 18 different genes (Table 2). More than one mutation was found in seven genes (*ANKRD11, CASK, EHMT1, GRIN2B, MECP2, SHANK3,* and *TRIO*); *SHANK3* resulted to be the most mutated gene. Among the identified variants, six were in X-linked genes and 12 in genes associated with autosomal dominant conditions. Twenty variants were absent from the gnomAD database, while five were known pathogenic mutations of the *MECP2, SATB2, SHANK3,* and *GRIN2B* genes.

Most of the mutations were *de novo*; one patient carried two of them. In 20 cases, paternity and maternity were established, while for two others, the parent DNA samples were not available. We can assume that the two truncating variants in *ANKRD11* and *MECP2* should be *de novo*,

since they involve genes that are associated with highly penetrant disorders. Furthermore, the p.Arg168X variant in the *MECP2* gene is a recurrent pathogenic variant associated with Rett syndrome. We also identified four inherited causative variants, one maternally inherited variant in the TRIO gene, which is associated with an autosomal dominant disorder, and three maternally inherited variants in X-linked genes (*ATRX, GRIA3,* and *RAB39B*). X-inactivation analysis was consistent with the phenotype expression. The *RAB39B* variant is thought to be also responsible for the mild phenotype reported in the mother, clinically re-evaluated after the molecular finding.

### *In silico* structural analysis may help predict mutation effects

15 causative variants (1 splicing, 6 frameshift, and 8 stop codon variants) are predicted to result in truncated proteins, if escaping the nonsense mediated mRNA decay. The *SHANK3* splicing variant has been shown to functionally impair mRNA splicing, producing an aberrant transcript containing an additional 77 bp intronic sequence (Li et al., 2018). Nine out of the 12 missense mutations were predicted pathogenic by the majority of computational methods, while pathogenicity predictions were discordant for three variants (Table 2). Among the nine variants with strong prediction of pathogenicity, p.Arg696His in *GRIN2B* has previously reported as pathogenic and has been shown to alter the Agonist Binding Domain (ABD) reducing channel activity (Swanger et al., 2016). I*n silico* analysis of two other missense mutations, *DYRK1A* p.Lys174Asn and *EHMT1* p.Gly1193Arg allowed us to classify them as loss of function mutations. Lysine 174 maps to the catalytic pocket of the *DYRK1A* kinase domain and alters the electrostatic surface of the domain, which is important for nucleotide binding (Figure 1). Glycine 1193 maps on the *EHMT1* SET domain, which is necessary for methylation of lysine-9 in the histone H3 N-terminus, and is buried in the rigid structure of the domain. A substitution of Glycine 1193 with an arginine residue should result in unfolding of the domain core (Figure 2).

We also hypothesized a predictive impact on the protein function of two variants with discordant pathogenicity: p.Phe193Leu in *RAB39B* (Supplementary Figure S2) and p.Tyr387His in *CASK* (Figure 3). Phenylalanine 193 maps in the RAB39B hypervariable C-terminal tail, which mediates interactions with effector proteins for proper intracellular targeting (Chavrier et al., 1991). This non-conservative substitution may disrupt a functional motif involved in protein interaction and cause a mislocalization of RAB39B, as shown for the protein mutated at position p.Gly192Arg closeby (Mata et al., 2015).

The Tyr387 residue maps on the N-terminal L27 domain, which mediates hetero- and homo-dimerization of the CASK protein. Comparing L27 domain sequence from different proteins, the 387 position is one of the most conserved and connected in the structure, denoting its important

structural role (Figure 3A and 3B). However, at that position histidine is more frequent than tyrosine, thus the p.Tyr387His should not have dramatically effect on the L27 domain structure (Supplementary Figure S3). However, when considering more related homologs of human CASK, the tyrosine 387 is highly conserved, presenting mostly tyrosine or, in few cases, phenylalanine compared to other L27 domains; this might indicate that tyrosine is important for the protein function and that a change to histidine might be deleterious, particularly in human and nearest species. We hypothesize that this substitution may result in a hypomorphic mutant protein with a reduced ability to form dimers (Figure 3), which is consistent with the phenotype of the boy, hemizygous for the *CASK* mutation. Indeed, individuals carrying hypomorphic mutations have been reported with an X-linked intellectual disability with or without nystagmus and additional clinical features (Hackett et al., 2010).

**Genotype – phenotype correlation**

One of the major criteria to assign the pathogenicity of the variants is the patient phenotype correlation with the classical syndrome associated to the corresponding gene. Most of the identified causative mutations (n=22) correlate with the previously reported phenotype of the corresponding gene (Table 2). For instance, the two known *MECP2* mutations were found in two girls with a suspected Rett syndrome. The p.Arg168* variant found in MR414.01 patient was missed in previous single gene testing since it was in a mosaic state. The p.Arg294* variant found in MR2145.01 patient was identified by Sanger sequencing of the *MECP2* region not covered by the 74 gene panel. A previous panel for Rett-Angelman spectrum disorders also missed the variant. In addition, for the two cases carrying a *de novo* mutation in *EHMT1* gene, the patient phenotypes were consistent with a Kleefstra Syndrome (KS). Both patients presented with core symptoms of the disease, including a moderate to severe intellectual disability with absent speech, and hypotonia. However, a Kleefstra syndrome was suspected for the characteristic carp-shaped mouth, only in MR2243.01 carrying the frameshift mutation, while MR2166.01 was originally suspected to have Smith Magenis Syndrome (SMS) probably due to brachycephaly, which is also a common characteristic of KS.

Nonetheless, in other cases the probands lacked some peculiar clinical features of the associated syndromes that prevented the geneticists to formulate hypothesis about a suspected syndrome. As an example, the two patients carrying a truncating mutation in the *ANKRD11* gene, both lacked, macrodontia of the upper central incisors, craniofacial features, and skeletal anomalies typical of the KBG syndrome, but both have mild ID, behavioral issues, and hearing loss (Sirmaci et al., 2011).

On the other hand, in some cases the detected mutations were found in genes never associated with particular phenotypic traits. In particular, we observed macrocephaly in three individuals carrying pathogenic mutations in three different genes (*ATRX*, *GRIN2B*, *and TRIO*) previously associated to microcephaly (Table 4). Interestingly, the *de novo* p.Arg696His in *GRIN2B* had been previously reported in a girl with a significant phenotypic overlap with our patient (MR 2019.01), including developmental delay, poor speech, intellectual disability, ASD with stereotypic behavior (Swanger et al., 2016). However, our patient present macrosomia and marked macrocephaly. There is only one report in literature associating *GRIN2B* with macrocephaly, in a case with an ~2 Mb interstitial deletion in 12p13 involving the entire *GRIN2B* gene, in addition to other genes (Morisada et al., 2016). Furthemore, we observed that the other two cases presenting with macrocephaly (MR 2276.01 and MR 984.01), carrying the maternally inherited *ATRX* and the *de novo TRIO* mutations, also carried other rare inherited CNVs or sequence variants (Supplementary Tables S3 and S4). Since inherited from healthy parents these alterations were classified as benign or of uncertain significance.

# Discussion

An accurate clinical and molecular diagnosis can greatly improve the treatment of individuals with neurodevelopmental disorders. However, differentiating between pathophysiological conditions that are clinically and genetically heterogeneous is a big challenge particularly when they arise with common comorbid disorders. The recent advent of Next Generation Sequencing (NGS) technologies allowed discovering many genes involved in these conditions. The study of undiagnosed cases with similar clinical manifestations through whole exome or genome sequencing highlighted the wide spectrum of phenotypic expression of some well-known genetic conditions, such as Rett syndrome, that can be caused by different genes involved in common biological pathways (Ehrhart et al., 2018). We are moving from a phenotypic-based to a gene-centered view of the diseases and we are incorporating the concept of disease network in the classification of the genetic conditions (Barabási et al., 2011). Due to the cost, ethical problems, and storage resources limitations, the use of genome and exome sequencing in the clinical practice is still a challenge. However, the availability of bench top systems, such as Ion Torrent platform, allowed spreading the use of gene panel sequencing in diagnostic laboratory for testing different genes involved in heterogeneous conditions, such as neurodevelopmental disorders.

In this study, we show the application of targeted sequencing of 74 genes in a cohort of 150 individuals with ID and/or ASD designed for diagnostic purpose. A certain causative variants

12

have been found in 17.3% of the tested individuals, the diagnostic yield is the same in the 93 individuals presenting ID and ASD comorbidity. In contrast to other studies, the majority of the diagnosed cases have a more severe ID (Redin et al., 2014) (Table 1). Considering patients presenting other comorbidities, such as epilepsy, microcephaly or macrocephaly, the diagnostic yield further increases to 20%, 31.6%, and 36.4%, respectively. Thus, the ability to find a molecular cause with this selection of genes is more effective in severe cases than those with a mild phenotype are. The use of network parameters to filter panel genes allowed selection of core genes, which are often hub genes implicated in multiple biological processes. Thus, an alteration of these genes may have consequences in diverse neurological systems leading to a more severe phenotype.

Interestingly, in three patients presenting macrocephaly as one of the clinical manifestations, we found a causative variant in three different genes (*ATRX*, *GRIN2B,* and *TRIO*) that have not been associated before to this feature. We decided to classify these variants as potentially causative even if the proband's phenotypes were only partially consistent with the phenotype reported for each gene. In these cases, to support the pathogenic role of the variants we considered the complex genetic architecture underlying the pathogenic mechanisms involved in NDDs (Woodbury-Smith and Scherer, 2018). The p.Arg696His in *GRIN2B* gene has been previously reported in an individual with ID but no macrocephaly, and experimentally tested for its ability to impair the protein function (Swanger et al., 2016). The other two cases carrying a *de novo TRIO* and a maternally inherited *ATRX* mutation were found to carry other rare inherited SNVs or CNVs. We can speculate that these rare inherited alterations may contribute to the disease as modifier variants interacting with causative ones to determine a specific phenotype. Diverse rare CNV, such as 15q11.2 and 16p12.1 deletions, have been implicated in multiple neurodevelopmental disorders, and a multi-hit model have been suggested (Girirajan et al., 2010; Abdelmoity et al., 2012). Furthermore, a multifactorial model has been proposed to explain the heritability of ASD, where rare *de novo* and inherited variations act within the context of a common-variant genetic load (Chaste et al., 2017; Guo et al., 2018)). This might also explain the variable clinical outcomes associated with the same causal variant, such as in the case of p.Arg696His in *GRIN2B.*

Another consideration to take into account is that for some NDD candidate genes a clear description of the related disorder will become possible only with the accumulation of reported cases. The *TRIO* gene has been recently associated with mild to borderline intellectual disability, delay in acquisition of motor and language skills, and neurobehavioral problems. Other findings can include microcephaly, variable digital and dental abnormalities, and suggestive facial features (Ba et al., 2016; Pengelly et al., 2016). Only few individuals carrying pathogenic

13

mutations in this gene have been reported to date, thus the clinical manifestations of the *TRIO*-related disorder is still evolving. Here, we report two novel individuals carrying pathogenic *TRIO* mutations, expanding the phenotypic spectrum associated with this novel NDDs gene. Furthermore, different mutations in the same gene can have different effects on the gene product, and therefore different pathological consequences (Barabási et al., 2011). These variants may perturbe specific subset of links in the interactome. For instance, different mutations in *ARX* gene, a paradigm of a pleiotropic gene, have been associated with diverse defects involving GABAergic neurons and associated with a wide spectrum of disorders (Friocourt and Parnavelas, 2010). This may also be the case for the p.His1371Tyr mutation in the GEF1 domain of *TRIO*. Recently, it has been show that variants affecting different protein interaction interfaces of this domain can produce bidirectional alterations of glutamatergic synapse function. The p.His1371Tyr maps far away from the Rac1 binding interface, near another variant, the p.Asp1368Val, that has been previously reported in a boy with severe ID (de Ligt et al., 2012). In contrast to other GEF1 domain variants involved in ASD, the p.Asp1368Val has been demonstrated to results in TRIO hyperfunction (Sadybekov et al., 2017). This finding may explain the more severe phenotype associated to the p.Asp1368Val, and those of the proband we report, carrying the p.His1371Tyr, presenting with severe ID, epilepsy, absent speech, and macrocephaly.

With the targeted approach, a proportion of patients remains without molecular diagnosis due to the limited number of sequenced genes. However, the higher coverage obtained by the gene panel sequencing, compared to a whole genome or exome sequencing approach, allows detecting a high number of variants in the target regions. Furthermore, due to the relatively small number of variants to be further investigate, this approach allows focusing on rare variants that would be filtered out for discordant pathogenicity predictions, as the novel hypomorphic *de novo CASK* mutation found in hemizygous state in a male with intellectual disability. For these variants, an in depth *in silico* analysis of the protein structure and function allowed to pinpoint possible mutation effects to support their pathogenic role. Finally, with this approach we provided a set of genotype-phenotype associations for a core set of genes involved in NDDs, which can be used to train or test computational methods for prioritization of potential disease-causing variants.

## CAGI-competition

At the completion of our study and before submitting for publication, genetic data and corresponding phenotypes of the tested individuals have been provided for a challenge at the Critical Assessment of Genome Interpretation (CAGI). CAGI is a worldwide blind test to assess computational methods for predicting phenotypic impacts of genomic variations. The outcome of

the predictive competition using the ID-ASD gene panel data set are presented elsewhere. However, here we provide the original data with few updates that can be used to train and/or to test their own computational tools/approach aiming to predict comorbid phenotypes from genetic variants in a subset of NDDs genes.

# Conclusions

The heterogeneity of NDDs reflects perturbations of the complex intra- and intercellular networks. The emerging tools of genomic medicine allows to hold the promise of disentangling the complex genetic architecture of these particular disorders, leading to the identification of different etiologies in similar phenotypes as well as common pathways underlying apparently distinct conditions. Based on the finding that diseases that share genes or involve proteins interacting with each other show elevated comorbidity, we designed a 74-gene panel to perform sequencing of individuals with ID and ASD comorbidities. With this approach, we identified disease causing variants or putative pathogenic variants in 27% of the tested individuals. This work demonstrates that knowledge about shared genes and common pathways can be used to develop innovative diagnostic tools needed to discriminate among overlapping phenotypes with high risk of developing comorbid features.

# Acknowledgments

# Disclosure statement

The authors declare no conflict of interest.

# References

1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. 2015. A global reference for human genetic variation. Nature 526:68–74.

Abdelmoity AT, LePichon J-B, Nyp SS, Soden SE, Daniel CA, Yu S. 2012. 15q11.2 proximal imbalances associated with a diverse array of neuropsychiatric disorders and mild dysmorphic features. J Dev Behav Pediatr 33:570–576.

Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet Chapter 7:Unit7.20.

Almuhtaseb S, Oppewal A, Hilgenkamp TIM. 2014. Gait characteristics in individuals with intellectual disabilities: a literature review. Res Dev Disabil 35:2858–2883.

Alva V, Nam S-Z, Söding J, Lupas AN. 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res 44:W410–5.

Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, et al. 2004. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 32:D115–9.

Barabási A-L, Gulbahce N, Loscalzo J. 2011. Network medicine: a network-based approach to human disease. Nat Rev Genet 12:56–68.

Ba W, Yan Y, Reijnders MRF, Schuurs-Hoeijmakers JHM, Feenstra I, Bongers EMHF, Bosch DGM, De Leeuw N, Pfundt R, Gilissen C, De Vries PF, Veltman JA, et al. 2016. TRIO loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. Hum Mol Genet 25:892–902.

Bettella E, Di Rosa G, Polli R, Leonardi E, Tortorella G, Sartori S, Murgia A. 2013. Early-onset epileptic encephalopathy in a girl carrying a truncating mutation of the ARX gene: rethinking the ARX phenotype in females. Clin Genet 84:82–85.

Blackburn PR, Tischer A, Zimmermann MT, Kemppainen JL, Sastry S, Knight Johnson AE, Cousin MA, Boczek NJ, Oliver G, Misra VK, Gavrilova RH, Lomberk G, et al. 2017. A Novel Kleefstra Syndrome-associated Variant That Affects the Conserved TPLX Motif within the Ankyrin Repeat of EHMT1 Leads to Abnormal Protein Folding. J Biol Chem 292:3866–3876.

Chaste P, Roeder K, Devlin B. 2017. The Yin and Yang of Autism Genetics: How Rare De Novo and Common Variations Affect Liability. Annu Rev Genomics Hum Genet 18:167–187.

Chavrier P, Gorvel JP, Stelzer E, Simons K, Gruenberg J, Zerial M. 1991. Hypervariable C-terminal domain of rab proteins acts as a targeting signal. Nature 353:769–772.

Chiurazzi P, Pirozzi F. 2016. Advances in understanding - genetic basis of intellectual disability. F1000Res 5.:

Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics 31:2745–2747.

Chun S, Fay JC. 2009. Identification of deleterious mutations within three human genomes.

Genome Res 19:1553–1561.

Colell EA, Iserte JA, Simonetti FL, Marino-Buslje C. 2018. MISTIC2: comprehensive server to study coevolution in protein families. Nucleic Acids Res 46:W323–W328.

Cristino AS, Williams SM, Hawi Z, An J-Y, Bellgrove MA, Schwartz CE, Costa L da F, Claudianos C. 2014. Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. Mol Psychiatry 19:294–301.

Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, Leyva N, Konidari I, Gentry RC, Hulme WF, Van Booven D, Mayo V, et al. 2014. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. Mol Autism 5:1.

Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. PLoS Comput Biol 6:e1001025.

Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res 37:e67.

Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, Liu X. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 24:2125–2137.

Ehrhart F, Sangani NB, Curfs LMG. 2018. Current developments in the genetics of Rett and Rett-like syndrome. Curr Opin Psychiatry 31:103–108.

Endele S, Rosenberger G, Geider K, Popp B, Tamer C, Stefanova I, Milh M, Kortüm F, Fritsch A, Pientka FK, Hellenbroich Y, Kalscheuer VM, et al. 2010. Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. Nat Genet 42:1021–1026.

Epi4K Consortium, Epilepsy Phenome/Genome Project, Allen AS, Berkovic SF, Cossette P, Delanty N, Dlugos D, Eichler EE, Epstein MP, Glauser T, Goldstein DB, Han Y, et al. 2013. De novo mutations in epileptic encephalopathies. Nature 501:217–221.

Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, Mering C von, Jensen LJ. 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808–15.

Friocourt G, Parnavelas JG. 2010. Mutations in ARX Result in Several Defects Involving GABAergic Neurons. Front Cell Neurosci 4:4.

Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, Georgieva L, Rees E, Palta P, Ruderfer DM, Carrera N, Humphreys I, et al. 2014. De novo mutations in schizophrenia implicate synaptic networks. Nature 506:179–184.

Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. 2009. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics 25:i54–62.

Giardina E, Peconi C, Cascella R, Sinibaldi C, Nardone AM, Novelli G. 2008. A multiplex molecular assay for the detection of uniparental disomy for human chromosome 15.

Electrophoresis 29:4775–4779.

Gibson TJ, Dinkel H, Van Roey K, Diella F. 2015. Experimental detection of short regulatory motifs in eukaryotic proteins: tips for good practice as well as for bad. Cell Commun Signal 13:42.

Gilman SR, Chang J, Xu B, Bawa TS, Gogos JA, Karayiorgou M, Vitkup D. 2012. Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. Nat Neurosci 15:1723–1728.

Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D. 2011. Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. Neuron 70:898–907.

Girirajan S, Rosenfeld JA, Cooper GM, Antonacci F, Siswara P, Itsara A, Vives L, Walsh T, McCarthy SE, Baker C, Mefford HC, Kidd JM, et al. 2010. A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. Nat Genet 42:203–209.

Guo H, Wang T, Wu H, Long M, Coe BP, Li H, Xun G, Ou J, Chen B, Duan G, Bai T, Zhao N, et al. 2018. Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. Mol Autism 9:64.

Hackett A, Tarpey PS, Licata A, Cox J, Whibley A, Boyle J, Rogers C, Grigg J, Partington M, Stevenson RE, Tolmie J, Yates JR, et al. 2010. CASK mutations are frequent in males and cause X-linked nystagmus and variable XLMR phenotypes. Eur J Hum Genet 18:544–552.

Harripaul R, Noor A, Ayub M, Vincent JB. 2017. The Use of Next-Generation Sequencing for Research and Diagnostics for Intellectual Disability. Cold Spring Harb Perspect Med 7.:

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee Y-H, Narzisi G, Leotta A, Kendall J, Grabowska E, et al. 2012. De novo gene disruptions in children on the autistic spectrum. Neuron 74:285–299.

Jagadeesh KA, Wenger AM, Berger MJ, Guturu H, Stenson PD, Cooper DN, Bernstein JA, Bejerano G. 2016. M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. Nat Genet 48:1581–1586.

Janson G, Zhang C, Prado MG, Paiardini A. 2017. PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. Bioinformatics 33:444–446.

Jehl P, Manguy J, Shields DC, Higgins DG, Davey NE. 2016. ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. Nucleic Acids Res 44:W11–5.

Jensen M, Girirajan S. 2017. Mapping a shared genetic basis for neurodevelopmental disorders. Genome Med 9:109.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.

Kazeminasab S, Najmabadi H, Kahrizi K. 2018. Intellectual Disability and Ataxia: Genetic Collisions. Arch Iran Med 21:29–40.

Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet 46:310–315.

Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE. 2017. Pathogenic variant burden in the ExAC database: an empirical approach to evaluating population data for clinical variant interpretation. Genome Med 9:13.

Krumm N, O'Roak BJ, Shendure J, Eichler EE. 2014. A de novo convergence of autism genetics and molecular neuroscience. Trends Neurosci 37:95–105.

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, et al. 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44:W90–7.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291.

Leonard H, Glasson E, Nassar N, Whitehouse A, Bebbington A, Bourke J, Jacoby P, Dixon G, Malacova E, Bower C, Stanley F. 2011. Autism and intellectual disability are differentially related to sociodemographic background at birth. PLoS One 6:e17875.

Ligt J de, Willemsen MH, Bon BWM van, Kleefstra T, Yntema HG, Kroes T, Vulto-van Silfhout AT, Koolen DA, Vries P de, Gilissen C, Rosario M del, Hoischen A, et al. 2012. Diagnostic exome sequencing in persons with severe intellectual disability. N Engl J Med 367:1921–1929.

Li Q, Wang K. 2017. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. Am J Hum Genet 100:267–280.

Li Y, Jia X, Wu H, Xun G, Ou J, Zhang Q, Li H, Bai T, Hu Z, Zou X, Xia K, Guo H. 2018. Genotype and phenotype correlations for SHANK3 de novo mutations in neurodevelopmental disorders. Am J Med Genet A 176:2668–2676.

Mata IF, Jang Y, Kim C-H, Hanna DS, Dorschner MO, Samii A, Agarwal P, Roberts JW, Klepitskaya O, Shprecher DR, Chung KA, Factor SA, et al. 2015. The RAB39B p.G192R mutation causes X-linked dominant Parkinson's disease. Mol Neurodegener 10:50.

Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S, Fraser MI, Gough J, Haft DR, et al. 2018. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res.

Morisada N, Ioroi T, Taniguchi-Ikeda M, Juan Ye M, Okamoto N, Yamamoto T, Iijima K. 2016. A 12p13 GRIN2B deletion is associated with developmental delay and macrocephaly. Hum Genome Var 3:16029.

Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, Lin C-F, Stevens C, Wang L-S, Makarov V, Polak P, Yoon S, et al. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485:242–245.

O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, Girirajan S, Karakoc E, Mackenzie AP, Ng SB, Baker C, Rieder MJ, Nickerson DA, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nat Genet 43:585–589.

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, Carvill G, Kumar A, Lee C, Ankenman K, Munson J, Hiatt JB, et al. 2012. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 338:1619–1622.

Paoletti P, Bellone C, Zhou Q. 2013. NMDA receptor subunit diversity: impact on receptor properties, synaptic plasticity and disease. Nat Rev Neurosci 14:383–400.

Pengelly RJ, Greville-Heygate S, Schmidt S, Seaby EG, Jabalameli MR, Mehta SG, Parker MJ, Goudie D, Fagotto-Kaufmann C, Mercer C, DDD Study, Debant A, et al. 2016. Mutations specific to the Rac-GEF domain of TRIO cause intellectual disability and microcephaly. J Med Genet 53:735–742.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, Thiruvahindrapuram B, Xu X, Ziman R, Wang Z, Vorstman JAS, Thompson A, et al. 2014. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet 94:677–694.

Piton A, Redin C, Mandel J-L. 2013. XLID-causing mutations and associated genes challenged in light of data from large-scale human exome sequencing. Am J Hum Genet 93:368–383.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res 20:110–121.

Potenza E, Di Domenico T, Walsh I, Tosatto SCE. 2015. MobiDB 2.0: an improved database of intrinsically disordered and mobile proteins. Nucleic Acids Res 43:D315–20.

Redin C, Gérard B, Lauer J, Herenger Y, Muller J, Quartier A, Masurel-Paulet A, Willems M, Lesca G, El-Chehadeh S, Le Gras S, Vicaire S, et al. 2014. Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. J Med Genet 51:724–736.

Reijnders MRF, Kousi M, Woerden GM van, Klein M, Bralten J, Mancini GMS, Essen T van, Proietti-Onori M, Smeets EEJ, Gastel M van, Stegmann APA, Stevens SJC, et al. 2017. Variation in a range of mTOR-related genes associates with intracranial volume and intellectual disability. Nat Commun 8:1052.

Reva B, Antipin Y, Sander C. 2011. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res 39:e118.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 17:405–424.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. Nat Biotechnol 29:24–26.

Rose PW, Prlić A, Altunkaya A, Bi C, Bradley AR, Christie CH, Costanzo LD, Duarte JM, Dutta S, Feng Z, Green RK, Goodsell DS, et al. 2017. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res 45:D271–D281.

Sadybekov A, Tian C, Arnesano C, Katritch V, Herring BE. 2017. An autism spectrum disorder-related de novo mutation hotspot discovered in the GEF1 domain of Trio. Nat Commun 8:601.

Schneider A, Dessimoz C, Gonnet GH. 2007. OMA Browser--exploring orthologous relations across 352 complete genomes. Bioinformatics 23:2180–2182.

Schwarz JM, Cooper DN, Schuelke M, Seelow D. 2014. MutationTaster2: mutation prediction for the deep-sequencing age. Nat Methods 11:361–362.

Shihab HA, Gough J, Cooper DN, Stenson PD, Barker GLA, Edwards KJ, Day INM, Gaunt TR. 2013. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat 34:57–65.

Shihab HA, Gough J, Mort M, Cooper DN, Day INM, Gaunt TR. 2014. Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. Hum Genomics 8:11.

Sim N-L, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res 40:W452–7.

Sirmaci A, Spiliopoulos M, Brancati F, Powell E, Duman D, Abrams A, Bademci G, Agolini E, Guo S, Konuk B, Kavaz A, Blanton S, et al. 2011. Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. Am J Hum Genet 89:289–294.

Swanger SA, Chen W, Wells G, Burger PB, Tankovic A, Bhattacharya S, Strong KL, Hu C, Kusumoto H, Zhang J, Adams DR, Millichap JJ, et al. 2016. Mechanistic Insight into NMDA Receptor Dysregulation by Rare Variants in the GluN2A and GluN2B Agonist Binding Domains. Am J Hum Genet 99:1261–1280.

Totsika V, Hastings RP, Emerson E, Lancaster GA, Berridge DM. 2011. A population-based investigation of behavioural and emotional problems and maternal mental health: associations with autism spectrum disorder and intellectual disability. J Child Psychol Psychiatry 52:91–99.

Tranchevent L-C, Ardeshirdavani A, ElShal S, Alcaide D, Aerts J, Auboeuf D, Moreau Y. 2016. Candidate gene prioritization with Endeavour. Nucleic Acids Res 44:W117–21.

Vissers LELM, Ligt J de, Gilissen C, Janssen I, Steehouwer M, Vries P de, Lier B van, Arts P, Wieskamp N, Rosario M del, Bon BWM van, Hoischen A, et al. 2010. A de novo paradigm for mental retardation. Nat Genet 42:1109–1112.

Woodbury-Smith M, Scherer SW. 2018. Progress in the genetics of autism spectrum disorder. Dev Med Child Neurol 60:445–451.

# Tables

| Features | Patients (n=150) | Yield with causative mutations | Yield with causative or putative mutations |
|---|---|---|---|
| **Gender** | | | |
| Female | 58 (39%) | 10 (17.2%) | 17 (29.3%) |
| Male | 92 (61%) | 16 (17.4%) | 24 (26%) |
| Age (at diagnosis) | 2- 42 y.o | 4-27 y.o. | 3-42 y.o. |
| **Familial History** | | | |
| Sporadic | 121 (81%) | 23 (19%) | 31(25.6%) |
| Familial | 29 (19%) | 3 (10.3%) | 10 (34.5%) |
| Sib pair | 5 (3.3%) | 0 | 0 |
| X-linked | 3 (2%) | 1 | 0 |
| **Intellectual Disability** | **146 (97.3%)** | **26 (17.8%)** | **41 (28%)** |
| Mild | 37 (24.6%) | 5 (13.5%) | 8 (21.6%) |
| Moderate | 38 (25.3%) | 8 (21%) | 10 (26.3%) |
| Severe | 33 (22%) | 9 (27.3%) | 12 (36.3%) |
| Not evaluated | 38 (25.3%) | 4 (10.5%) | 11 (28.9%) |
| **Comorbidity** | | | |
| ASD (autistic features) | 93 (62%) | 15 (16.1%) | 25 (26.8%) |
| ASD not reported | 16 (10.6%) | 3 (18.7%) | 5 (31.2%) |
| Epilepsy | 55 (36.6%) | 11 (20%) | 17 (30.9%) |
| Hypotonia | 28 (18.6%) | 6 (21.4%) | 6 (21.4%) |
| Ataxia | 11 (7.3%) | 2 (18%) | 3 (27.2%) |
| Microcephaly | 19 (12%) | 6 (31.5%) | 8 (42.1%) |
| Macrocephaly | 11 (7.3%) | 4 (36.4%) | 4 (36.4%) |
| **Previous investigation** | | | |
| aCGH | 125 (83.3%) | 22 (17.6%) | 35 (28%) |
| Fragile-X | 86 (57.3%) | 10 (11.6%) | 18 (20.9%) |

22

| | | | |
|---|---|---|---|
| EEG anomaly | 56 (37.3%) | 9 (16%) | 17 (30.4%) |
| MRI anomaly | 37 (24.6%) | 9 (24.3%) | 13 (35.1%) |
| Other tests | 57 (38%) | 13 (22.8%) | 18 (31.6%) |

**Table 1:** Description of the cohort of 150 individuals enrolled for the study of ID/ASD comorbidity.

| Patient | Sex | Gene | Mutation | Mode of inheritance | Variant segregation | dbSNP | gnomAD AC/AN | Prediction consensus | CADD | Correlation with classic phenotype | InterVar manually adjusted (criteria) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2033.01 | M | *ANKRD11* | chr16:89345974:CCTTCGGGG:C; NM.013275; c.6968_6975del; p.Ala2323GlyfsX205 | AD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PM2, PM4, PM6, PP4) |
| 2338.01 | F | *ANKRD11* | chr16:89346137:A:G; NM.013275; c.6812_6813delCT; p.Pro2271ArgfsX23 | AD | _ | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PM2, PP4) |
| 2127.01 | F | *ARID1B* | chr6:157528165:G:T; NM.017519; c.5851G>T; p.Glu1951X | AD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PM2, PM6, PP3, PP4) |
| 2276.01 | M | *ATRX* | chrX:76909661:T:C; NM.138270; c.4244A>G; p.Asn1415Ser | XLR | Maternal, Balanced 40% | rs782562458 | 1/177812 | 9/12 | 23 | Partially | Likely pathogenic (PS2, PM2, PP3, PP6) |
| 602.01 | F | *CASK* | chrX:41401980:G:A; NM_001126055; c.2032C>T:p.Q678X | XLD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PS2, PV1, PM2, PP3) |
| 196.01 | M | *CASK* | chrX:41448842:A:G; NM_001126055; c.1141T>C; p.Tyr381His | XLD | *de novo* | _ | _ | 5/12 | 20.8 | Partially | Likely pathogenic (PM1, PM2, PP3) |

| 2222.01 | M | *DYRK1A* | chr21:38858777:G:C; NM.001369; c.525G>C; p.Lys175Asn | AD | *de novo* | _ | _ | 8/12 | 25.8 | Yes | Pathogenic (PS2, PM1, PM2, PP3, PP4) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2166.01 | M | *EHMT1* | chr9:140728837:G:C; NM.024757; c.3577G>C; p.Gly1193Arg | AD | *de novo* | _ | _ | 11/12 | 34 | Yes † | Likely pathogenic (PM1, PM2, PM6, PP3, BP1) |
| 2243.01 | M | *EHMT1* | chr9:140657209:GA:G; NM.024757; c.1585delA; p.Ser529ValfsX34 | AD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PS2, PM2, PM4, PP4) |
| 2140.01 | M | *GRIA3* | chrX:122460015:G:A;NM.000828; c.647G>A; p.Arg216Gln | XLR | Maternal affected, Skewed 26% | rs753214982 | 8/199967 | 7/12 | 23.3 | Yes | Likely pathogenic (PM1, PM2, PP3, PP6) |
| 2278.01 | M | *GRIN2B* | chr12:13761626:T:G;NM.000834; c.1921A>C; p.Ile641Leu | AD | *de novo* | _ | _ | 7/12 | 28.6 | Yes | Likely pathogenic (PM1, PM2, PM6) |
| 2019.01 | M | *GRIN2B* | chr12:13724822:C:T;NM.000834; c.2087G>A; p.Arg696His | AD | *de novo* | _ | _ | 6/12 | 35 | Yes ‡ | Pathogenic (PS2, PM1, PM2, PP3, PP6) |
| 2145.01 | F | *MECP2* | chrX:53296399:C:T; NM.004992; c.880C>T; p.Arg294X | XLD | *de novo* | rs61751362 | 3/183432 (1 Hem) | _ | _ | Yes | Pathogenic (PVS1, PS2, PS4, PM2, PP2, PP3, PP4, PP5) |
| 414.01 | F | *MECP2* | chrX:153296777:G:A; NM.004992; c.502C>T; p.Arg168X | XLD | _ | rs61748421 | _ | _ | _ | Yes | Pathogenic (PVS1, PS4, PM2, PP3, PP4, PP5) |

24

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1730.01 | F | *OPHN1* | chrX:67273488:C:T; NM.002547; c.2323G>A; p.Val775Met | XLR | *de novo* | _ | _ | 6/12 | 24.3 | Yes | Likely pathogenic (PM2, PM6, PP3, PP4) |
| 1974.01 | M | *RAB39B* | chrX:154490151:A:C; NM.171998; c.579T>G; p.Phe193Leu | XLR | maternal affected, skewed 66% | rs782042596 | 2/183440 (2 Hem) | 3/12 | 8.86 | Partially | Uncertain significance (PM1, PM2, PP4) |
| 1985.01 | M | *SATB2* | chr2:200213882:G:A; NM.001172509; c.715C>T; p.Arg239X | AD | *de novo* | rs137853127 | _ | _ | _ | Yes | Pathogenic (PVS1, PS2, PM2, PP3, PP5) |
| 2274.01 | M | *SETBP1* | chr18:42531499:AGAGC:-; NM.015559; c.2194.2198del; p.Glu734AlafsX18 | AD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PS2, PS4, PP3) |
| 1970.01 | F | *SHANK3* | chr22:51159830:A:TTC; NM.033517.1; c.[3568_3569insTT;3569A>C]; p.Asp1190ValfsX5 | AD | *de novo* | _ | _ | _ | _ | Yes § | Likely pathogenic (PM2, PM4, PM6, PP4) |
| 2230.01 | F | *SHANK3* | chr22:51153476:G:A; NM.033517.1; c.2265+1G>A | AD | *de novo* | _ | 1/158210 | _ | _ | Yes | Pathogenic (PVS1, PM2, PM6, PP3, PP4) |
| 2271.01 | M | *SHANK3* | chr22:51159718:C:T; NM.033517.1 c.3457C>T;p.Arg1153X | AD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PM2, PM6, PP4) |
| 1749.01 | M | *SHANK3* | chr22:51160432:GA:G; NM.033517.1; c.4172delA; p.Glu1391AspfsX36 | AD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PS2, PM2, PP4) |
| 2233.01 | M | *SYNGAP1* | chr6:33411228:C:T; NM.006772; c.2899C>T; p.Arg967X | AD | *de novo* | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PS2, PM1, PM2, |

25

| Patient | Sex | Gene | Mutation | Mode of inheritance | Variant segregation | dbSNP | gnomAD AC/AN | Prediction consensus | CADD | Correlation with classic phenotype | InterVar manually adjusted (criteria) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | 26 | | | | | | PP3, PP4) |
| 984.01 | M | TRIO | chr5:14390392:C:T; NM.007118; c.4111C>T; p.His1371Tyr | AD | de novo | _ | _ | 7/12 | 27 | Partially | Likely pathogenic (PS2, PM1, PM2, PP3) |
| 2165.01 | M | TRIO | chr5:14394159:C:T; NM.007118; c.4231C>T; p.Arg1411X | AD | maternal, affected | _ | _ | _ | _ | Yes | Pathogenic (PVS1, PM2, PP3, PP4) |
| 2113.01 | M | MED13L | chr12:116445337:C:T; NM.015335; c.2117G>A; p.Gly706Glu | AD | de novo | rs200257416 | 7/251400 | 5/12 | 20.3 | Yes | Likely pathogenic (PS2, PM7, PP4) |
| | | ASH1L | chr1:155449342:T:C; NM.018489; c.3319A>G; p.Ile1107Val | AD | de novo | rs140137038 | 148/282442 | 1/12 | 0.6 | Partially | Uncertain significance (PS2, BP4) |

**Table 2: Causative variants found in the 150 patient cohort.** For maternally inherited variants, based on X-inactivation analysis, the quote of expressed mutated allele in the mother was reported. Prediction consensus is calculated among the twelve methods provided by ANNOVAR (see Methods section). Variant interpretation based on InterVar: criteria manually adjusted based on our findings (Richards et al., 2015). M, male; F, female; Genomic position on human GRCh37; AD: autosomal dominant; AR: autosomal recessive; XLD: X-linked dominant; XLR: X-linked recessive; † according to the *EHMT1* phenotypic spectrum described in (Blackburn et al., 2017); ‡ This *GRIN2B* missense mutation has been previously found in a female with a phenotype similar to our case (Swanger et al., 2016). § This *SHANK3* splicing mutation has been previously reported by (Li et al., 2018).

| Patient | Sex | Gene | Mutation | Mode of inheritance | Variant segregation | dbSNP | gnomAD AC/AN | Prediction consensus | CADD | Correlation with classic phenotype | InterVar manually adjusted (criteria) |
|---|---|---|---|---|---|---|---|---|---|---|---|

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 2264.01 | M | *ANKRD11* | chr16:89349967:T:C; NM_013275; c.2983A>G; p.Lys995Glu | AD | NA | _ | | 6/12 | 24 | Partially | Uncertain significance (PM2, PP3, BP1) |
| 2322.01 | M | *ATRX* | chrX:76764055:T:A; NM_138270; c.7253A>T; p.Tyr2418Phe | XLD/XLR | _ | _ | _ | 9/12 | 22.7 | Yes | Uncertain significance (PM2, PP3, PP4) |
| 2141.01 | M | *CHD8* | chr14:21876977:G:A; NM_001170629 c.2372C>T; p.Pro791Leu | AD | Not in mother | rs372717272 | 2/236228 | 12/12 | 32 | Partially | Likely benign (PM1, PP3, BS2, BP1) |
| 2389.01 | M | *CHD8* | chr14:21882498:T:C; NM_001170629; c.1267A>G; p.Lys423Glu | AD | _ | _ | _ | 7/12 | 27.4 | Partially | Uncertain significance (PM1, PM2, BP1) |
| 2374.01 | F | *CNTNAP2* | chr7:148080864:C:T; NM_014141; c.3599C>T; p.Ser1200Leu | AR/AD | _ | rs778312206 | 3/251396 | 8/12 | 23.3 | Yes | Uncertain significance (PM1, PM2, PP4, BP1) |
| 2039.01 | M | *CREBBP* | chr16:3788561:C:T; NM_004380; c.4393G>A; p.Gly1465Arg | AD | _ | _ | _ | 11/12 | 26.6 | Yes | Likely pathogenic (PM1, PM2, PP3, PP4) |
| 2375.01 | F | *DEAF1* | chr11:684897:C:T; NM_001293634; c.730+1G>A | AR, AD | _ | _ | _ | _ | _ | Partially | Likely pathogenic (PVS1, PM2) |
| 243.01 | F | *GRIN2B* | chr12:13720096:C:G; NM_000834; c.2461G>C; p.Val821Leu | AD | _ | _ | _ | 7/12 | 28.8 | Partially | Uncertain significance (PM1, PM2, PP2) |
| 1975.01 | M | *KATNAL2* | chr18:44595922:C:T; NM_031303; c.743C>T; p.Ala248Val | AD | maternal, low penetrance | rs140833601 | 57/277052 | 11/12 | 34 | Partially | Uncertain significance |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | (PM1, PP3) |
| 2344.01 | M | PHF8 | chrX:53964467:A:G; NM_001184897; c.2794T>C; p.Cys932Arg | XLR | maternal, Skewed 70% | rs782094119 | 3/112150 (2 Hem) | 1/12 | 2.49 | No | Uncertain significance (PM2, PM7, BP4) |
| 2272.01 | M | PTEN | chr10:89690828:G:A; NM_000314; c.235G>A; p.Ala79Thr | AD | maternal | rs202004587 | 29/281404 | 7/12 | 22 | Partially | Uncertain significance (PM1, PM2, PP3) |
| 2340.01 | M | SCN2A | chr2:166165900:C:T; NM_001040142; c.644C>T; p.Ala215Val | AD | maternal | rs149024364 | 1/246128 | 10/12 | 14.97 | Partially | Uncertain significance (PM1, PM3, PM6) |
| 2007.01 | M | SHANK2 | chr11:70644598:G:A; NM_012309; c.1727C>T; p.Pro576Gln | AD | _ | _ | _ | 10/12 | 32 | No | Uncertain significance (PM2, PP3) |
| 1769.01 | M | FOXP1 | chr3:71026867:A:C; NM_001244810; c.1355T>G; p.Ile452Ser | AD | paternal | _ | _ | 7/12 | 21.7 | Partially | Uncertain significance (PM1, PM2, PP3) |
| | | CNTNAP2 | chr7:146829502:G:T; NM_014141; c.1249G>T:p.Asp417Tyr | AR/AD | maternal | rs147815978 | 9/245852 | 7/12 | 20.5 | Partially | Uncertain significance (PM1, PM2, PP3, BP1) |
| 2053.01 | F | GAD1 | chr2:171678594:T:C; NM_000817; c.83-3T>C | AR | maternal | rs769951300 | 1/234994 | _ | _ | Partially | _ |
| | | GAD1 | chr2:171702114:C:T; NM_000817; c.850C>T; p.Leu284Phe | AR | paternal | rs780519382 | 3/251444 | 7/12 | 30 | | Likely pathogenic (PM1, PM2, PM3, PP2, PP3) |

**Table 3 Putative pathogenic variants found in the 150 patients cohort.** For maternally inherited variants, the quote of expressed mutated allele in the mother, based on X-inactivation analysis, was reported. Prediction consensus is calculated among the twelve methods provided by ANNOVAR (see Methods section). Variant interpretation based on InterVar: criteria manually adjusted based on our findings (Richards et al., 2015).M, male; F, female; Genomic position on human GRCh37; AD: autosomal dominant; AR: autosomal recessive; XLD: X-linked dominant; XLR: X-linked recessive; Hem: hemizygote.

| Clinical features | N° individuals affected (Tot=150) | Genes carrying causative variants | Genes carrying likely pathogenic variants |
|---|---|---|---|
| Intellectual disability | | | |
| Autistic traits | 73 | ANKRD11, ARID1B, CASK, EHMT1, GRIA3, GRIN2B, MECP2, MED13L (ASH1L), SHANK3, SYNGAP1 | ANKRD11, CHD8, FOXP1 (CNTNAP2), DEAF1, GRIN2B, KATNAL2, PHF8, PTEN, SCN2A |
| Epilepsy | 40 | ANKRD11, EHMT1, GRIA3, MECP2, OPHN1, RAB39B, SETBP1, SYNGAP1, TRIO | ANKRD11, CASK, CHD8, CREBBP, GAD1, SCN2A, SHANK2 |
| Microcephaly | 19 | ARID1B, CASK, DYRK1A, MECP2, TRIO | **CHD8,** CREBBP, **SHANK2** |
| Macrocephaly | 12 | **ATRX**, GRIA3, **GRIN2B**, **TRIO** | - |
| Hypotonia | 28 | CASK, DYRK1A, EHMT1, MECP2, **TRIO** | - |
| Ataxia | 11 | MECP2 | GAD1 |

**Table 4: Mutated genes in the different phenotypic manifestations (ASD, epilepsy, Microcephaly, Macrocephaly, Hypotonia, and Ataxia).** Some genes have been found mutated in individuals presenting phenotypic traits that have not been previously associated to these genes (highlighted in red, bold).
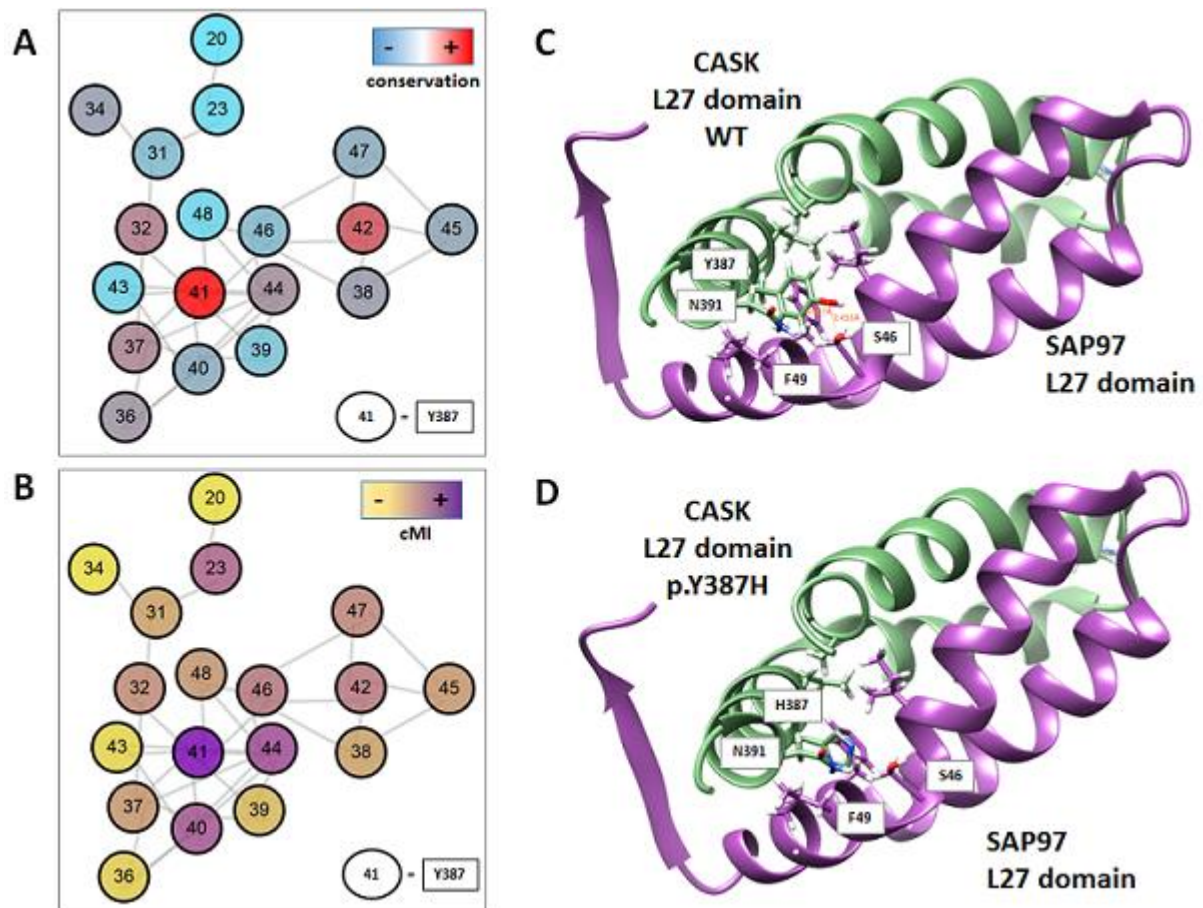
# Figures
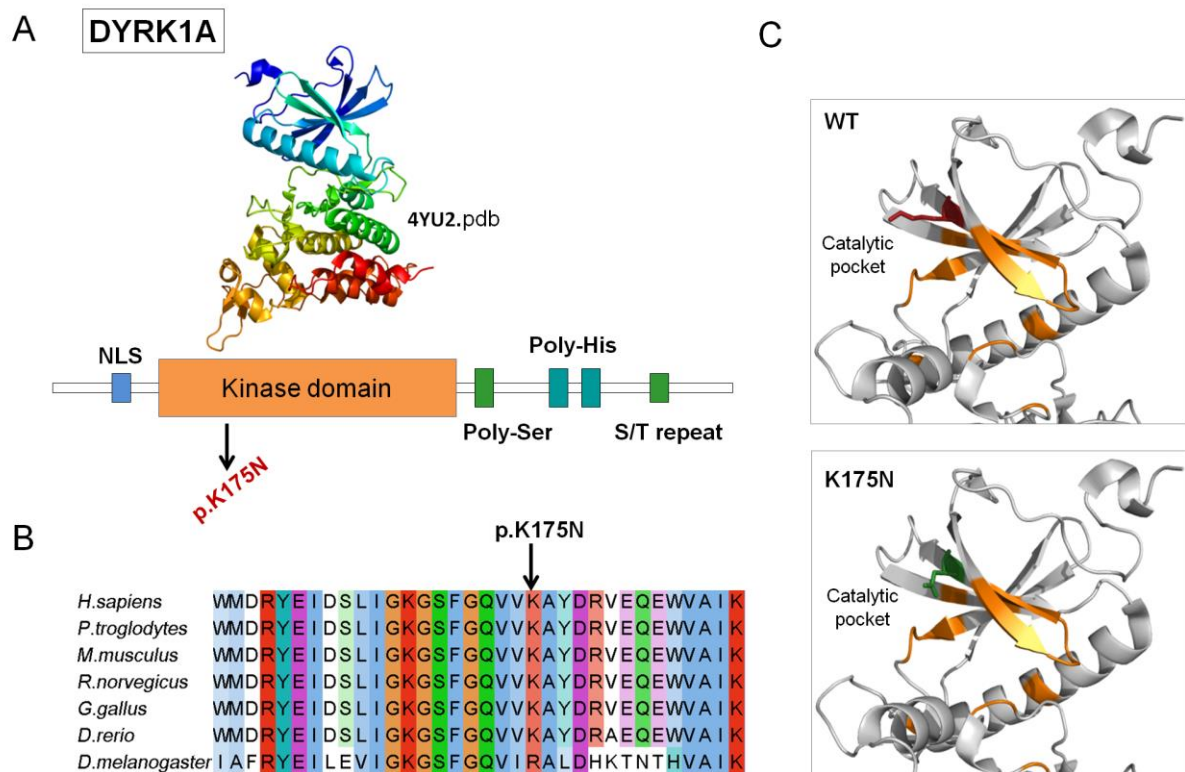


**Figure 1. DYRK1A missense mutation affect catalytic pocket of kinase domain**

**A)** Domain architecture of DYRK1A. Protein sequence presents regions biased toward polar (serine and threonine) and aromatic (histidine residues). NLS = nuclear localization signal. **B)** DYRK1A p.Lys175 and neighboring residues are conserved among orthologous sequences. Amino acids are colored by conservation, according to ClustalX color code. **C)** DYRK1A p. Lys175Asn variant and wild type residues are mapped to kinase domain structure (4yu2.pdb, chain A). Residues involved in nucleotide binding are represented in orange sticks, wild-type lysine (K) in red and asparagine (N) in green.
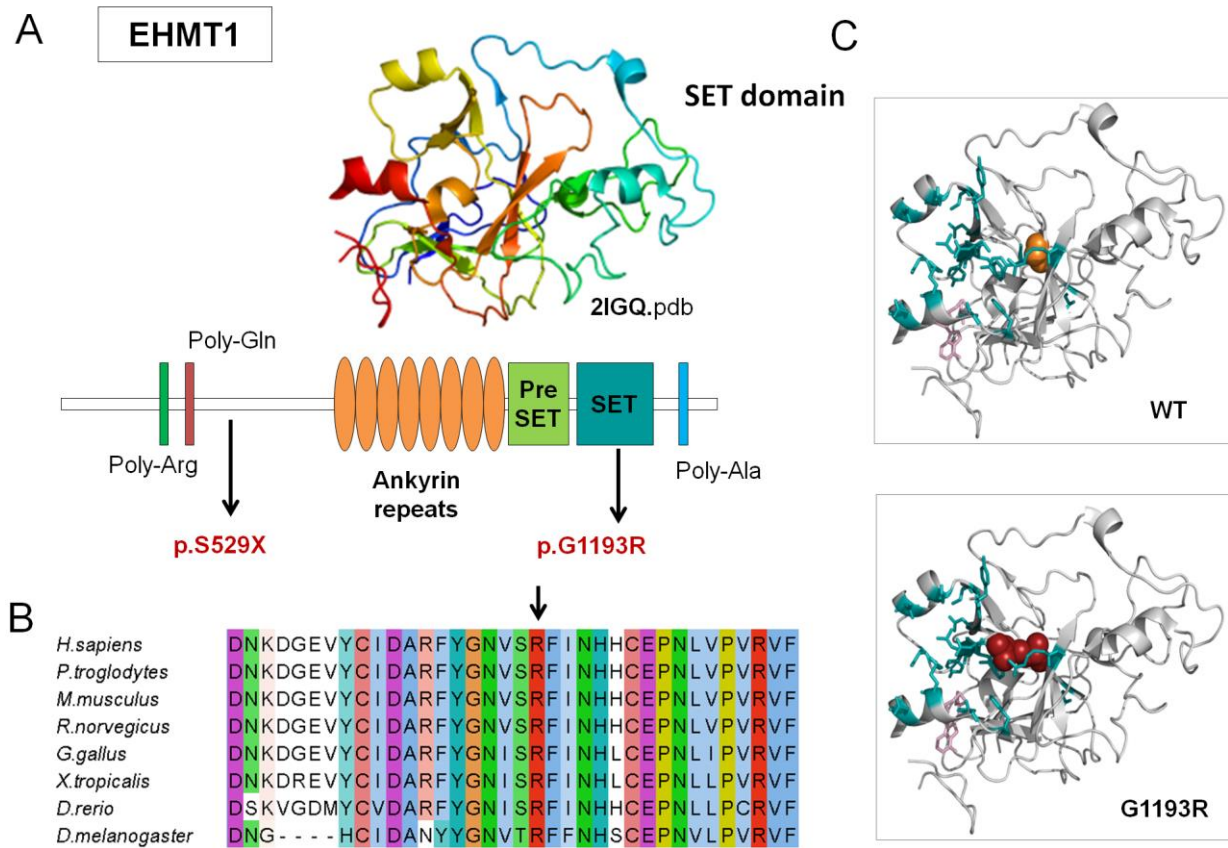
**Figure 2. EHMT1 missense mutation affects SET domain.**

Causative variants identified in our patient cohort are mapped to the EHMT1. Protein sequence presents regions biased toward polar (glutamine and arginine) and a poly-alanine motif. The ankyrin domain (orange) is involved with the histone H3K9me binding. The Pre-SET domain (green) contributes to SET domain stabilization **B)** EHMT1 p.Gly1193 and neighboring residues are conserved among orthologous sequences. Amino acids are colored by conservation, according ClustalX color code. **C)** EHMT1 p.Gly1193Arg variant (red) and wild type glycine (orange) are mapped to SET domain structure (2igq.pdb, chain A) Residues involved in H3K9 binding and the S-adenosyl-L-methionine molecule are represented in blue sticks.

**Figure 3. The novel likely hypomorphic mutation of CASK may alter L27 domain dimerization**

A) Covariation network of L27 domain (PF02828). Nodes are the residues (L27 domain numbering) coloured by conservation from red to light blue (highest to lowest respectively). B) Nodes are coloured by cumulative Mutual information (cMI) from violet to yellow (highest to lowest respectively). Edges are the top 0.1% covariation scores (Mutual Information) calculated with Mistic2 (Colell et al., 2018). C) Ribbon representation of the L27- SAP97 complex. CASK domain L27 is coloured green (chain B 1RSO PDB) and SAP97 violet (chain A 1RSO PDB). In sticks are shown the residues that interact (their R) with Y387 in the complex. D) same colouring schema. Y387 was "*in silico*" mutated to H387. Mutation and figures were generated by UCSF Chimera (Pettersen et al., 2004).