ThermoFisher
SCIENTIFIC

A VIRTUAL EVENT

5 DAYS
OF STEM CELLS

Connect. Discover. Advance.
Join us for the world's leading virtual stem cell event.

gibco

Stay on the leading edge of stem cell research

The Gibco™ 5 Days of Stem Cells virtual event connects you to the latest stem cell techniques, research breakthroughs, and esteemed scientists from around the world–all from the comfort of anywhere

The 5-day virtual agenda is packed full of incredible insights in the form of:

- Leading scientific presentations from thought leaders around the world

- Behind-the-scenes virtual training demos

- Scientific poster sessions

- Hundreds of key stem cell tools and resources

- A global network of researchers including our stem cell experts and technical support

We'll share developments, discoveries, and cutting-edge content connected to a wide variety of stem cell applications, including disease modeling, cell and gene therapy, 3D modeling, and much more.

It's all happening October 12–16, 2020.

CURRENT PROTOCOLS
A Wiley Brand

Register for free at **thermofisher.com/5daysofstemcells**

ThermoFisher
SCIENTIFIC

SPECIAL ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Characterization of intellectual disability and autism comorbidity through gene panel sequencing

Maria C. 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, Department of Woman and Child Health, University of Padova, C.so Stati Uniti, 4, Padova, Italy

[2]Fondazione Istituto di Ricerca Pediatrica, Città della Speranza, Padova, Italy

[3]Department of Biomedical Sciences, University of Padova, Padova, Italy

[4]Medical Genetics Unit, Ospedale Universitario S. Anna, Ferrara, Italy

[5]Medical Genetics Unit, San Martino Hospital, Belluno, Italy

[6]Epilepsy and Child Neurophysiology Unit, Scientific Institute IRCCS E. Medea, Treviso, Italy

[7]Medical Genetics Unit, Dolo General Hospital, Venezia, Italy

[8]Fondazione IRCCS, Ca Granda, Pediatric Highly Intensive Care Unit, Milan, Italy

[9]Child Neuropsychiatry Unit, Epilepsy Center, Department of Health Sciences, Santi Paolo-Carlo Hospital, University of Milano, Milano, Italy

[10]Division of Medical Genetics, Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, Utah

[11]Paediatric Neurology Unit, Department 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, Treviso, Italy

[14]Genetic Counseling Service, Department of Pediatrics, Regional Hospital of Bolzano, Bolzano, Italy

[15]Bioinformatics Unit, Fundación Instituto Leloir, Buenos Aires, Argentina

[16]Institute of Neuroscience, National Research Council, Padova, Italy

Correspondence
Emanuela Leonardi, Molecular Genetics of Neurodevelopment, Department of Woman and Child Health, University of Padova, 35129, Padova, Italy.
Email: emanuela.leonardi@unipd.it

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 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 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

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

1 | INTRODUCTION

Neurodevelopmental 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% and 3% in general population, depending on patient and parent age, or the measure of intellectual quotient used (Leonard et al., 2011). The ID is characterized by deficits in both intellectual and adaptive functioning that first manifest during early childhood. Children with ID exhibit increased risk to present potential co-occurring developmental conditions, such as autism spectrum disorders (ASD; 28%), epilepsy (22.2%), stereotypic movement disorders (25%), and motor disorders, which substantially affect daily living and well-being (Almuhtaseb, Oppewal, & Hilgenkamp, 2014; Jensen & Girirajan, 2017; Kazeminasab, Najmabadi, & Kahrizi, 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 an ID alone (Totsika, Hastings, Emerson, Lancaster, & Berridge, 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 a 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 (Epi4K Consortium, Epilepsy Phenome/Genome Project et al., 2013; Fromer et al., 2014; Neale et al., 2012). It has been estimated that mutations in more than 1,000 different genes might cause ID (Chiurazzi & 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 (Cukier et al., 2014; Vissers et al., 2010). 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, Noor, Ayub, & Vincent, 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 (Krumm, O'Roak, Shendure, & Eichler, 2014; Vissers et al., 2010). 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, that is 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 ID, epilepsy, and schizophrenia (Cristino et al., 2014; Endele et al., 2010; Gilman et al., 2011; Krumm et al., 2014; Paoletti, Bellone, & Zhou, 2013; 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, Gulbahce, & Loscalzo, 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 ASD 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-comparative genomic hybridization (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.

2 | MATERIALS AND METHODS

2.1 | Patient selection

Patients were referred from clinical geneticists of 17 Italian public hospitals with a diagnosis of nonspecific 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, while Table S1 reports for each of 150 patients the presence of ID, ASD, epilepsy, microcephaly or macrocephaly, hypotonia, and ataxia. 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.

2.2 | 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), OMIM, and PubMed. Candidate genes were extracted in particular from recent exome sequencing and meta-analysis studies (Table S2). 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 a training set for Endeavor 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 with ID or ASD only (i.e., genes with at least five evidence for ID or ASD; 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 (Table S3).

2.3 | Gene panel sequencing

Nucleic acids were extracted from blood samples using the Wizard genomic DNA Promega Kit (Promega Corporation). Multiplex, polymerase chain reaction (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).

TABLE 1 Description of the cohort of 150 individuals enrolled for the study of ID/ASD comorbidity

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 (year old; at diagnosis)	2–42	4–27	3–42
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%)
X-Fragile	86 (57.3%)	10 (11.6%)	18 (20.9%)
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%)

Abbreviations: ASD: autism spectrum disorder; aCGH: array-comparative genomic hybridization; EEG: electroencephalogram; MRI: magnetic resonance imaging.

2.4 | 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, that is 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 (Genomes Project Consortium et al., 1000, 2015), and ExAC (Kobayashi et al., 2017) databases, as well as in our in house database of 150 patients. We excluded single nucleotide variants (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, Redin, & Mandel, 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 12 computational methods. ANNOVAR provides predictions from SIFT (Sim et al., 2012), the Polyphen-2 (Adzhubei, Jordan, & Sunyaev, 2013) HDIV and HVAR versions, LRT (Chun & Fay, 2009), Mutation Taster (Schwarz, Cooper, Schuelke, & Seelow, 2014), MutationAssessor (Reva, Antipin, & Sander, 2011), FATHMM (Shihab et al., 2014); PROVEAN (Choi & 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, Hubisz, Rosenbloom, & Siepel, 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.

2.5 | 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 reverse-transcription polymerase chain reaction was performed using random primers. cDNA was used as a template in nested PCR reactions with specific primers to amplify the regions containing the mutation. PCR products were tested on 1.5% agarose gel and sequenced.

2.6 | Variant classification

A clinical interpretation of selected variants was first evaluated using InterVar (Li & Wang, 2017), providing an automated variant interpretation based on 18 criteria published by the American College of Medical Genetics and Genomics (ACMG). The InterVar web server has a manual adjustment step that allows reviewing the automated interpretation by selecting appropriate criteria according to additional information and knowledge about the involved domain.

According to the ACMG recommendations, we classified filtered candidate variants into five categories (pathogenic, likely pathogenic, uncertain significance, likely benign, benign) based on multiple lines of evidence (conservation, allele frequency in population databases, computational inferences of variant effect, mode of inheritance, X-inactivation pattern, and disease segregation; Figure 1). Assuming that a patient phenotype is consistent with a Mendelian disorder, variants be classified as pathogenic are considered causative, that is, responsible for the phenotypic manifestations of the carrier patient. Likely pathogenic variants instead require further investigation to be classified as pathogenic/causative variants, for example, segregation analysis and/or functional analysis. Rare or novel variants predicted as pathogenic and altering genes conferring an increased risk of autism have been classified as possible contributing factors if inherited from parents reported as healthy, as they alone are not sufficient to cause the disease (Table S5). The criteria used to classify the variants (Richards et al., 2015) are reported for both causative and likely pathogenic variants (Table 2 and 3). All the causative and likely pathogenic variants have been submitted to the LOVD database.

2.7 | 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, Dessimoz, & Gonnet, 2007) and aligned with MAFFT (Katoh & Standley, 2013). When available, crystal structures were retrieved from PDB (Rose et al., 2017). Structures of protein domains were modeled with MODELLER (Alva, Nam, Söding, & Lupas, 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, Iserte, Simonetti, & Marino-Buslje, 2018) to evaluate covariation between residues. Structures were manually explored with Pymol (Janson, Zhang, Prado, & Paiardini, 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, Di Domenico, Walsh, & Tosatto, 2015) and ELM (Gibson, Dinkel,
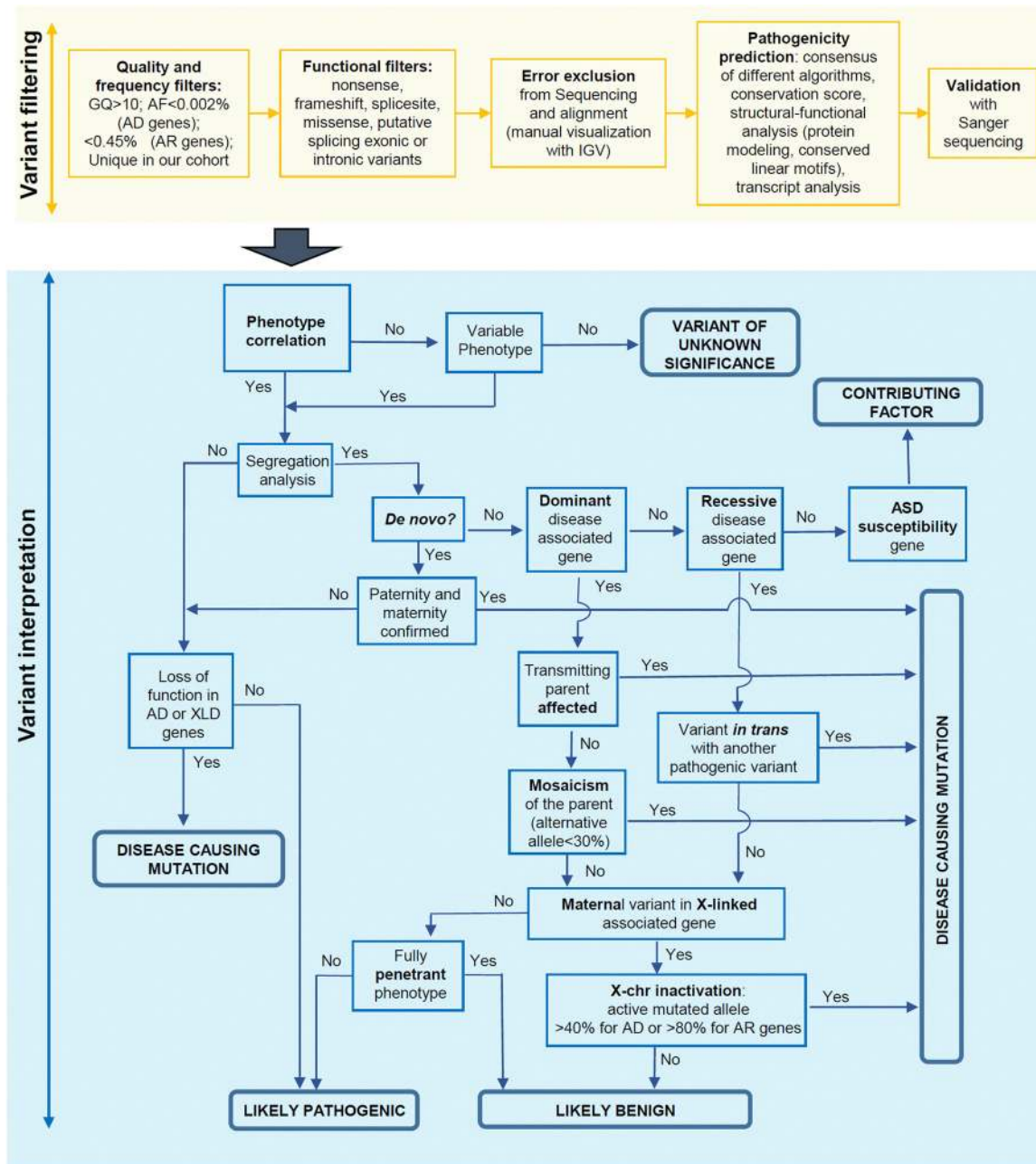
FIGURE 1 Workflow describing the important steps in the classification of selected variants. AD: autosomal dominant; AF: allele frequency; AR: autosomal recessive; GQ: genotype quality; XLD: X-linked, dominant

Van Roey, & Diella, 2015), using the interactive exploration tool ProViz (Jehl, Manguy, Shields, Higgins, & Davey, 2016).

3 | RESULTS

3.1 | 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 dominant 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 nonspecific 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

TABLE 2 Causative variants found in the 150 patient cohort

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	chr16:89345974CCTTCGGGG>C; NM_013275.5:c.6968_6975del; p.Ala2323Glyfs*206	AD	De novo	-	-	-	-	Yes	Pathogenic (PVS1, PM2, PM4, PM6, PP4)
2338.01	F	ANKRD11	chr16:89346136CAG>C; NM_013275.5:c.6812_6813del; p.Pro2271Argfs*24	AD	-	-	-	-	-	Yes	Pathogenic (PVS1, PM2, PP4)
2127.01	F	ARID1B	chr6:157528165G>T; NM_017519.2:c.5851G>T; p.Glu1951*	AD	De novo	-	-	-	-	Yes	Pathogenic (PVS1, PM2, PM6, PP3, PP4)
2276.01	M	ATRX	chrX:76909661T>C; NM_000489.4:c.4244A>G; p.Asn1415Ser	XLR	Maternal, random 60%	rs78256245-8	1/177812	9/12	23	Partially	Likely pathogenic (PS2, PM2, PP3, PP6)
602.01	F	CASK	chrX:41401980G>A; NM_003688.3:c.2119C>T; p.Gln707*	XLD	De novo	-	-	-	-	Yes	Pathogenic (PS2, PV1, PM2, PP3)
196.01	M	CASK	chrX:41448842A>G; NM_003688.3:c.1159T>C; p.Tyr387His	XLD	De novo	-	-	5/12	20.8	Partially	Likely pathogenic (PM1, PM2, PP3)
2222.01	M	DYRK1A	chr21:38858777G>C; NM_001396.4:c.525G>C; p.Lys175Asn	AD	De novo	-	-	8/12	25.8	Yes	Pathogenic (PS2, PM1, PM2, PP3, PP4)
2166.01	M	EHMT1	chr9:140728837G>C; NM_024757.4:c.3577G>C; p.Gly1193Arg	AD	De novo	-	-	11/12	34	Yes[a]	Likely pathogenic (PM1, PM2, PM6, PP3, BP1)
2243.01	M	EHMT1	chr9:140657209GA>G; NM_024757.4:c.1585del; p.Ser529Valfs*34	AD	De novo	-	-	-	-	Yes	Pathogenic (PVS1, PS2, PM2, PM4, PP4)
2140.01	M	GRIA3	chrX:122460015G>A; NM_000828.4:c.647G>A; p.Arg216Gln	XLR	Maternal affected, skewed 74%	rs753214982	8/199967	7/12	23.3	Yes	Likely pathogenic (PM1, PM2, PP3, PP6)
2278.01	M	GRIN2B	chr12:13761626T>G; NM_000834.4:c.1921A>C; p.Ile641Leu	AD	De novo	-	-	7/12	28.6	Yes	Likely pathogenic

(Continues)

TABLE 2 (Continued)

Patient	Sex	Gene	Mutation	Mode of inheritance	Variant segregation	dbSNP	gnomAD AC/AN	Prediction consensus	CADD	Correlation with classic phenotype	InterVar manually adjusted (criteria)
2019.01	M	GRIN2B	chr12:13724822C>T; NM_000834.4:c.2087G>A; p.Arg696His	AD	De novo	-	-	6/12	35	Yes[b]	Pathogenic (PS2, PM1, PM2, PP3, PP6)
2145.01	F	MECP2	chrX:153296399G>A; NM_004992.3:c.880C>T; p.Arg294*	XLD	De novo	rs61751362	3/183432 (1 Hem)	-	-	Yes	Pathogenic (PVS1, PS2, PS4, PM2, PP2, PP3, PP4, PP5)
414.01	F	MECP2	chrX:153296777G>A; NM_004992.3:c.502C>T; p.Arg168*	XLD	-	rs61748421	-	-	-	Yes	Pathogenic (PVS1, PS4, PM2, PP3, PP4, PP5)
1730.01	F	OPHN1	chrX:67273488C>T; NM_002547.2:c.2323G>A; p.Val775Met	XLR	De novo	-	-	6/12	24.3	Yes	Likely pathogenic (PM2, PM6, PP3, PP4)
1974.01	M	RAB39B	chrX:154490151A>C; NM_171998.2:c.579T>G; p.Phe193Leu	XLR	maternal affected, random 34%	r-s78204259-6	2/183440 (2 Hem)	3/12	8.86	Partially	Uncertain significance (PM1, PM2, PP4)
1985.01	M	SATB2	chr2:200213882G>A; NM_001172509.1:c.715C>T; p.Arg239*	AD	De novo	r-s13785312-7	-	-	-	Yes	Pathogenic (PVS1, PS2, PM2, PP3, PP5)
2274.01	M	SETBP1	chr18:42531498AAGAGC>A; NM_015559.2:c.2199_2203del; p.Glu734Alafs*18	AD	De novo	-	-	-	-	Yes	Pathogenic (PVS1, PS2, PS4, PP3)
1970.01	F	SHANK3	chr22:51159830A>TTC; NM_033517.1:c.[3568_3569insTT;3569A>C]; p.Asp1190Valfs*5	AD	De novo	-	-	-	-	Yes	Likely pathogenic (PM2, PM4, PM6, PP4)
2230.01	F	SHANK3	chr22:51153476G>A; NM_033517.1:c.2265+1G>A	AD	De novo	-	1/158210	-	-	Yes[c]	Pathogenic (PVS1, PM2, PM6, PP3, PP4)
2271.01	M	SHANK3	chr22:51159718C>T; NM_033517.1:c.3457C>T; p.Arg1153*	AD	De novo	-	-	-	-	Yes	Pathogenic (PVS1, PM2, PM6, PP3, PP4)

(Continues)

TABLE 2 (Continued)

Patient	Sex	Gene	Mutation	Mode of inheritance	Variant segregation	dbSNP	gnomAD AC/AN	Prediction consensus	CADD	Correlation with classic phenotype	InterVar manually adjusted (criteria)
1749.01	M	SHANK3	chr22:51160432GA>G; NM_033517.1:c.4172delA; p.Glu1391Aspfs*36	AD	De novo	–	–	–	–	Yes	Pathogenic (PVS1, PM2, PM6, PP4)
2233.01	M	SYNGAP1	chr6:33411228C>T; NM_006772.2:c.2899C>T; p.Arg967*	AD	De novo	–	–	–	–	Yes	Pathogenic (PVS1, PS2, PM2, PP4)
984.01	M	TRIO	chr5:14390392C>T; NM_007118.2:c.4111C>T; p.His1371Tyr	AD	De novo	–	–	7/12	27	Partially	Pathogenic (PVS1, PS2, PM1, PM2, PP3, PP4)
2165.01	M	TRIO	chr5:14394159C>T; NM_007118.2:c.4231C>T; p.Arg1411*	AD	Maternal, affected	–	–	–	–	Yes	Likely pathogenic (PS2, PM1, PM2, PP3)
2113.01	M	MED13L	chr12:116445337C>T; NM_015335.4:c.2117G>A; p.Gly706Glu	AD	De novo	rs200025741-6	7/251400	5/12	20.3	Yes	Likely pathogenic (PS2, PM7, PP4)
		ASH1L	chr1:155449342T>C; NM_018489.2:c.3319A>G; p.Ile1107Val	AD	De novo	rs14013703-8	148/282442	1/12	0.6	Partially	Uncertain significance (PS2, BP4)

Note: 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 12 methods provided by ANNOVAR (see Materials and Methods section). Variant interpretation based on InterVar: criteria manually adjusted based on our findings (Richards et al., 2015).

Abbreviations: AD: autosomal dominant; AR: autosomal recessive; F: female; GRCh37: genomic position on human; M: male; XLD: X-linked dominant; XLR: X-linked recessive.

[a] According to the *EHMT1* phenotypic spectrum described in (Blackburn et al., 2017).

[b] This *GRIN2B* missense mutation has been previously found in a female with a phenotype similar to our case (Swanger et al., 2016).

[c] This *SHANK3* splicing mutation has been previously reported by (Li et al., 2018). *SHANK3* variants named according to the *SHANK3* RefSeq mRNA (NM_033517.1) and protein (NP_277052.1) sequence, in which the exon 11 sequence has been corrected.

TABLE 3 Putative pathogenic variants found in the 150 patients cohort

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:g.89349967T>C; NM_013275.4; c.2983A>G; p.Lys995Glu	AD	NA	-	-	6/12	24	Partially	Uncertain significance (PM2, PP3, BP1)
2322.01	M	ATRX	chrX:g.76764055T>A; NM_000489.3; c.7253A>T; p.Tyr2418Phe	XLD/XLR	-	-	-	9/12	22.7	Yes	Uncertain significance (PM2, PP3, PP4)
2141.01	M	CHD8	chr14:g.21876977G>A; NM_001170629.1: c.2372C>T; p.Pro791Leu	AD	Not in mother	rs37271727-2	2/236228	12/12	32	Partially	Uncertain significance (PM1, PP3, BP1)
2389.01	M	CHD8	chr14:g.21882498T>C; NM_001170629.1: c.2104A>G; p.Lys702Glu	AD	-	-	-	7/12	27.4	Partially	Uncertain significance (PM1, PM2, BP1)
2374.01	F	CNTNAP2	chr7:g.148080864C>T; NM_014141.4; c.3599C>T; p.Ser1200Leu	AR/AD	-	rs77831220-6	3/251396	8/12	23.3	Yes	Uncertain significance (PM1, PM2, PP4, BP1)
2039.01	M	CREBBP	chr16:g.3788561C>T; NM_004380.2; c.4393G>A; p.Gly1465Arg	AD	-	-	-	11/12, splicing alteration	26.6	Yes	Likely pathogenic (PM1, PM2, PP3, PP4)
2375.01	F	DEAF1	chr11:g.684897C>T; NM_021008.2; c.870+1G>A	AR, AD	-	-	-	Splicing alteration	-	Partially	Likely pathogenic (PVS1, PM2)
243.01	F	GRIN2B	chr12:g.13720096C>G; NM_000834.3; c.2461G>C; p.Val821Leu	AD	-	-	-	7/12	28.8	Partially	Uncertain significance (PM1, PM2, PP2)
1975.01	M	KATNAL2	chr18:g.44595922C>T; NM_031303.1; c.743C>T; p.Ala248Val	AD	Maternal, low penetrance	rs14083360-1	57/277052	11/12	34	Partially	Uncertain significance (PM1, PP3)
2344.01	M	PHF8	chrX:g.53964467A>G; NM_001184897.1: c.2794T>C; p.Cys932Arg	XLR	Maternal, Skewed 30%	rs78209411-9	3/112150 (2 Hem)	1/12	2.49	No	Uncertain significance (PM2, PM7, BP4)
2272.01	M	PTEN	chr10:g.89690828G>A; NM_000314.4; c.235G>A; p.Ala79Thr	AD	Maternal	rs20200458-7	29/281404	7/12	22	Partially	Uncertain significance (PM1, PM2, PP3)
2340.01	M	SCN2A	chr2:g.166165900C>T; NM_001040142.1: c.644C>T; p.Ala215Val	AD	Maternal	rs144902436-4	1/246128	10/12	14.97	Partially	Uncertain significance (PM1, PM3, PM6)
2007.01	M	SHANK2	chr11:g.70644598G>A; NM_012309.3; c.1727C>T; p.Pro576Leu	AD	-	-	-	10/12	32	Partially	Uncertain significance (PM2, PP3)

(Continues)

TABLE 3 (Continued)

Patient	Sex	Gene	Mutation	Mode of inheritance	Variant segregation	dbSNP	gnomAD AC/AN	Prediction consensus	CADD	Correlation with classic phenotype	InterVar manually adjusted (criteria)
1769.01	M	FOXP1	chr3:g.71026867A>C; NM_032682.5; c.1355T>G; p.Ile452Ser	AD	Paternal	–	–	7/12	21.7	Partially	Uncertain significance (PM1, PM2, PP3)
		CNTNAP2	chr7:g.146829502G>T; NM_014141.4; 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:g.171678594T>C; NM_000817.2; c.83−3T>C	AR	Maternal	rs769951300	1/234994	–	–	Partially	–
		GAD1	chr2:g.171702114C>T; NM_000817.2; c.850C>T; p.Leu284Phe	AR	Paternal	rs780519382	3/251444	7/12	30		Likely pathogenic (PM1, PM2, PM3, PP2, PP3)

Note: 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 12 methods provided by ANNOVAR (see Materials and Methods section). Variant interpretation based on InterVar: criteria manually adjusted based on our findings (Richards et al., 2015).
Abbreviations: AD: autosomal dominant; AR: autosomal recessive; F: female; GRCh37: genomic position on human; Hem: hemizygote; M: male ; XLD: X-linked dominant; XLR: X-linked recessive.

meanwhile reclassified and the disease association confirmation is pending (*HDAC4* and *KIRREL3*; Table S3).

3.2 | 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 (<20×) 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 <10×. Depending on phenotype manifestations, otherwise, these regions have been covered by Sanger sequencing, as they could not be analyzed reliably.

3.3 | Cohort description and diagnostic yield

Our cohort of 150 individuals is enriched in males (61%). Sporadic cases of 81.3% 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%). Ninety-three patients (62%) had both ID and ASD. Fifty-three (35.3%) had the only ID, while four (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). Thirty-nine (26%) patients with ID and ASD present also epilepsy (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, structural pathogenic alteration of chromosomes was excluded by FISH, karyotype, or aCGH; however, in 17 of these, the aCGH analysis revealed a CNV inherited from unaffected parents or involving gene-poor regions (Table S4). Eighty-six 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.

3.4 | Variant detection and prioritization

In coding exons or exon-intron boundaries regions of the 74 genes, we detected on average about 74 SNVs, with a range of 60–80 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 Materials and Methods section, we selected about 170 variants for further analysis, 47 of which were absent from the general population (Table S5). 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.870+1G>A in *DEAF1*, p.Val821Leu in *GRIN2B*, and 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 are possible, only with functional assays supporting the pathogenic effect of the identified variants. In addition to a novel missense mutation in the *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 the *GAD1* gene, which is associated with an autosomal recessive phenotype. The girl presents with clinical features consistent with the few reported *GAD1* cases, for example, severe developmental delay and nonprogressive ataxia. 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 support their pathogenic role (Table S5).

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 the very mild phenotype (e.g. *TRIO* and *SLC6A1*; Table S5).

3.5 | Disease-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, 6 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, 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.

3.6 | In silico structural analysis may help predict mutation effects

Fifteen causative variants (one splicing, six frameshift, and eight 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 the strong prediction of pathogenicity, p.Arg696His in *GRIN2B* has previously reported as pathogenic and has been shown to alter the Agonist Binding Domain reducing channel activity (Swanger et al., 2016). In 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 2). 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 the unfolding of the domain core (Figure 3).

We also hypothesized a predictive impact on the protein function of two variants with discordant pathogenicity: p.Phe193Leu in *RAB39B* (Figure S2) and p.Tyr387His in *CASK* (Figure 4). 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 nonconservative 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 close by (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 4a,b). However, at that position histidine is more frequent than tyrosine, thus the p.Tyr387-His should not have a dramatic effect on the L27 domain structure (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 4), 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).

3.7 | 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 with 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 the *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 the severe ID 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 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 a 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 a 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 with 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, ID, ASD with stereotypic behavior (Swanger et al., 2016). However, our patient presents 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). Furthermore, 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 (Tables S4 and S5). Since inherited from healthy parents these alterations were classified as benign or of uncertain significance.

4 | 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 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, Sangani, & Curfs, 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
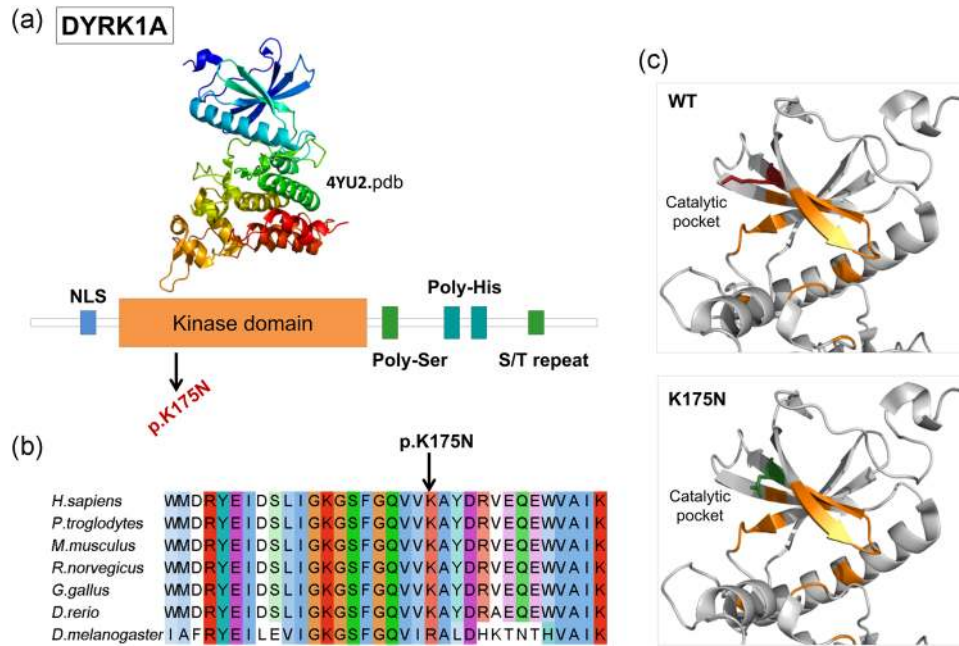
FIGURE 2 *DYRK1A* missense mutation affects the catalytic pocket of the kinase domain. (a) Domain architecture of DYRK1A. Protein sequence presents regions biased toward polar (serine and threonine) and aromatic (histidine residues). (b) DYRK1A p.Lys175 and neighboring residues are conserved among orthologous 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. NLS : nuclear localization signal
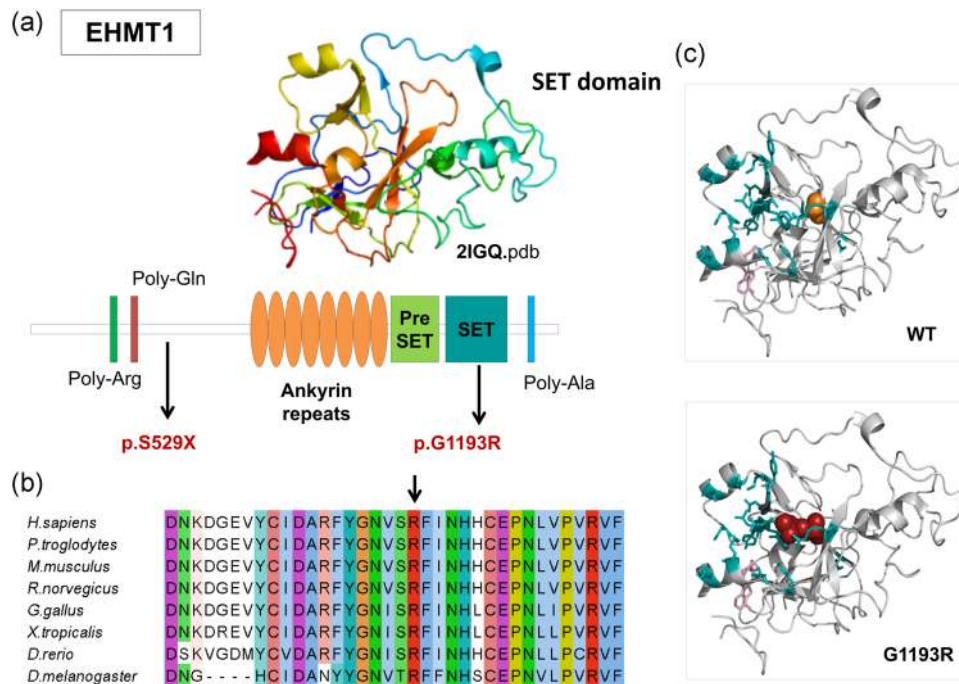
FIGURE 3 *EHMT1* missense mutation affects the SET domain. Causative variants identified in our patient cohort are mapped to the EHMT1. Protein sequence presents regions biased toward polar (glutamine and arginine) and a poly-alanine motif. The ankyrin domain (orange) is involved with the histone H3K9me 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 to 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
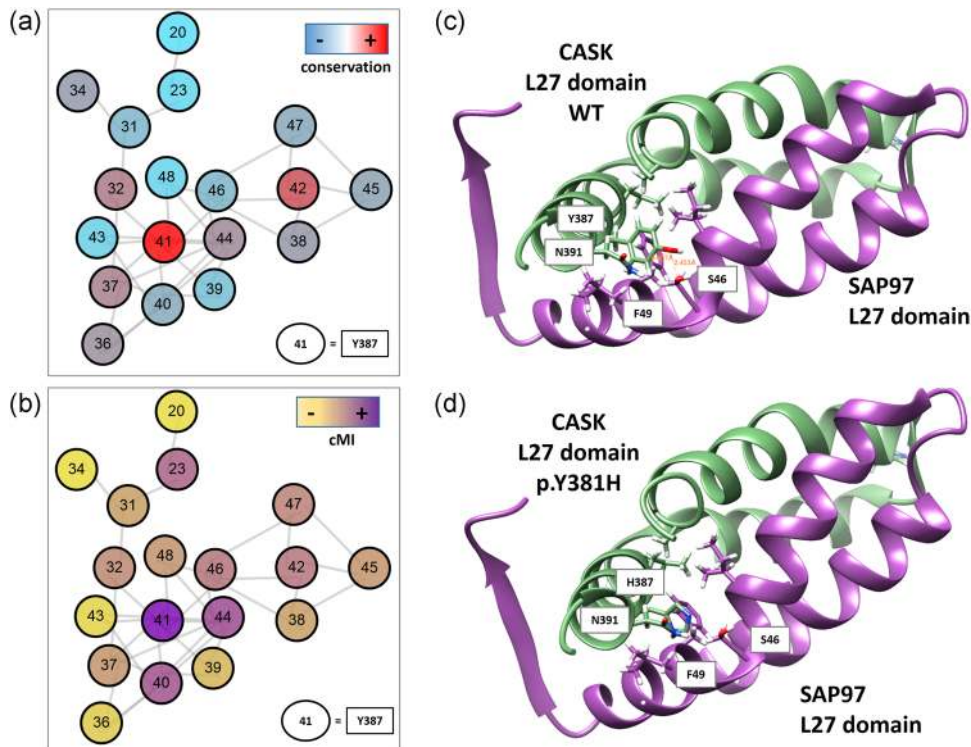
FIGURE 4 The novel likely hypomorphic mutation of *CASK* may alter L27 domain dimerization. (a) Covariation network of L27 domain (PF02828). Nodes are the residues (L27 domain numbering) colored by conservation from red to light blue (highest to lowest respectively). (b) Nodes are colored by cumulative 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 colored 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 coloring schema. Y387 was "in silico" mutated to H387. Mutation and figures were generated by UCSF Chimera (Pettersen et al., 2004)

use of genome and exome sequencing in the clinical practice is still a challenge. However, the availability of benchtop systems, such as Ion Torrent platform, allowed spreading the use of gene panel sequencing in the 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. Certain causative variants 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 & 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 has been suggested (Abdelmoity et al., 2012; Girirajan et al., 2010). 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, Roeder, & Devlin, 2017; Guo et al., 2018). This might also explain the variable clinical outcomes

TABLE 4 Mutated genes in the different phenotypic manifestations (ASD, epilepsy, microcephaly, macrocephaly, hypotonia, and ataxia)

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*

Note: Some genes have been found mutated in individuals presenting phenotypic traits that have not been previously associated to these genes (highlighted in bold).

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 ID, 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 a few individuals carrying pathogenic mutations in this gene have been reported to date, thus the clinical manifestations of the *TRIO*-related disorder are 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 & Parnavelas, 2010). This may also be the case for the p.His1371Tyr mutation in the GEF1 domain of TRIO. Recently, it has been shown 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, Tian, Arnesano, Katritch, & Herring, 2017). This finding may explain the more severe phenotype associated with 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 investigated, 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 the hemizygous state in a male with ID. 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.

4.1 | 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 is 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.

5 | CONCLUSIONS

The heterogeneity of NDDs reflects perturbations of the complex intracellular and intercellular networks. The emerging tools of genomic medicine allow 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 gene or involves 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 a high risk of developing comorbid features.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

ORCID

Maria C. Aspromonte http://orcid.org/0000-0002-4937-6952
Cristina Marino-Buslje http://orcid.org/0000-0002-6564-1920
Silvio C.E. Tosatto http://orcid.org/0000-0003-4525-7793
Emanuela Leonardi http://orcid.org/0000-0001-8486-8461

REFERENCES

Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., … Abecasis, G. R. 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature*, 526, 68–74.

Abdelmoity, A. T., LePichon, J.-B., Nyp, S. S., Soden, S. E., Daniel, C. A., & Yu, S. (2012). 15q11.2 proximal imbalances associated with a diverse array of neuropsychiatric disorders and mild dysmorphic features. *Journal of Developmental and Behavioral Pediatrics (Baltimore, MD)*, 33, 570–576.

Adzhubei, I, Jordan, DM, & Sunyaev, SR (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*, 76(1), 1–52. https://doi.org/10.1002/0471142905.hg0720s76

Almuhtaseb, S., Oppewal, A., & Hilgenkamp, T. I. M. (2014). Gait characteristics in individuals with intellectual disabilities: A literature review. *Research in Developmental Disabilities*, 35, 2858–2883.

Alva, V., Nam, S.-Z., Söding, J., & Lupas, A. N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Research*, 44, W410–W415.

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., … Natale, D. A. (2004). UniProt: The universal protein knowledgebase. *Nucleic Acids Research*, 32, D115–D119.

Ba, W., Yan, Y., Reijnders, M. R., Schuurs-Hoeijmakers, J. H., Feenstra, I., Bongers, E. M., … Veltman, J. A. (2016). TRIO loss of function is associated with mild intellectual disability and affects dendritic branching and synapse function. *Human Molecular Genetics*, 25, 892–902.

Barabási, A.-L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews Genetics*, 12, 56–68.

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. *Clinical Genetics*, 84, 82–85.

Blackburn, P. R., Tischer, A., Zimmermann, M. T., Kemppainen, J. L., Sastry, S., Knight Johnson, A. E., … Lomberk, G. (2017). A novel Kleefstra syndrome-associated variant that affects the conserved TPLX motif within the ankyrin repeat of EHMT1 leads to abnormal protein folding. *Journal of Biological Chemistry*, 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. *Annual Review of Genomics and Human Genetics*, 18, 167–187.

Chavrier, P., Gorvel, J. P., 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(599), 1–16. https://doi.org/10.12688/f1000research.7134.1

Choi, Y., & Chan, A. P. (2015). PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, 31, 2745–2747.

Chun, S., & Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Research*, 19, 1553–1561.

Colell, E. A., Iserte, J. A., Simonetti, F. L., & Marino-Buslje, C. (2018). MISTIC2: Comprehensive server to study coevolution in protein families. *Nucleic Acids Research*, 46, W323–W328.

Cristino, A. S., Williams, S. M., Hawi, Z., An, J.-Y., Bellgrove, M. A., Schwartz, C. E., … Claudianos, C. (2014). Neurodevelopmental and neuropsychiatric disorders represent an interconnected molecular system. *Molecular Psychiatry*, 19, 294–301.

Cukier, H. N., Dueker, N. D., Slifer, S. H., Lee, J. M., Whitehead, P. L., Lalanne, E., … Mayo, V. (2014). Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. *Molecular Autism*, 5, 1.

Davydov, E. V., Goode, D. L., Sirota, M., Cooper, G. M., Sidow, A., & Batzoglou, S. (2010). Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLOS Computational Biology*, 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 Research*, 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. *Human Molecular Genetics*, 24, 2125–2137.

Ehrhart, F., Sangani, N. B., & Curfs, L. M. G. (2018). Current developments in the genetics of Rett and Rett-like syndrome. *Current Opinion in Psychiatry*, 31, 103–108.

Endele, S., Rosenberger, G., Geider, K., Popp, B., Tamer, C., Stefanova, I., ... Kalscheuer, V. M. (2010). Mutations in GRIN2A and GRIN2B encoding regulatory subunits of NMDA receptors cause variable neurodevelopmental phenotypes. *Nature Genetics*, *42*, 1021–1026.

Epi4K Consortium, Epilepsy Phenome/Genome Project, Allen, A. S., Berkovic, S. F., Cossette, P., Delanty, N., Dlugos, D., ... Han, Y. (2013). De novo mutations in epileptic encephalopathies. *Nature*, *501*, 217–221.

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., ... von, Jensen, L. J. (2013). STRING v9.1: Protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Research*, *41*, D808–D815.

Friocourt, G., & Parnavelas, J. G. (2010). Mutations in ARX result in several defects involving GABAergic neurons. *Frontiers in Cellular Neuroscience*, *4*, 4.

Fromer, M., Pocklington, A. J., Kavanagh, D. H., Williams, H. J., Dwyer, S., Gormley, P., ... Humphreys, I. (2014). De novo mutations in schizophrenia implicate synaptic networks. *Nature*, *506*, 179–184.

Garber, M., Guttman, M., Clamp, M., Zody, M. C., Friedman, N., & Xie, X. (2009). Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics*, *25*, i54–i62.

Giardina, E., Peconi, C., Cascella, R., Sinibaldi, C., Nardone, A. M., & Novelli, G. (2008). A multiplex molecular assay for the detection of uniparental disomy for human chromosome 15. *Electrophoresis*, *29*, 4775–4779.

Gibson, T. J., 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 Communication and Signaling*, *13*, 42.

Gilman, S. R., Chang, J., Xu, B., Bawa, T. S., Gogos, J. A., Karayiorgou, M., & Vitkup, D. (2012). Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nature Neuroscience*, *15*, 1723–1728.

Gilman, S. R., 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, J. A., Cooper, G. M., Antonacci, F., Siswara, P., Itsara, A., ... Kidd, J. M. (2010). A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nature Genetics*, *42*, 203–209.

Guo, H., Wang, T., Wu, H., Long, M., Coe, B. P., Li, H., ... Zhao, N. (2018). Inherited and multiple de novo mutations in autism/developmental delay risk genes suggest a multifactorial model. *Molecular Autism*, *9*, 64.

Hackett, A., Tarpey, P. S., Licata, A., Cox, J., Whibley, A., Boyle, J., ... Yates, J. R. (2010). CASK mutations are frequent in males and cause X-linked nystagmus and variable XLMR phenotypes. *European Journal of Human Genetics*, *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 Harbour Perspectives in Medicine*, *7*(3), a026864.

Iossifov, I., Ronemus, M., Levy, D., Wang, Z., Hakker, I., Rosenbaum, J., ... Grabowska, E. (2012). De novo gene disruptions in children on the autistic spectrum. *Neuron*, *74*, 285–299.

Jagadeesh, K. A., Wenger, A. M., Berger, M. J., Guturu, H., Stenson, P. D., Cooper, D. N., ... Bejerano, G. (2016). M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nature Genetics*, *48*, 1581–1586.

Janson, G., Zhang, C., Prado, M. G., & 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, D. C., Higgins, D. G., & Davey, N. E. (2016). ProViz-a web-based visualization tool to investigate the functional and evolutionary features of protein sequences. *Nucleic Acids Research*, *44*, W11–W15.

Jensen, M., & Girirajan, S. (2017). Mapping a shared genetic basis for neurodevelopmental disorders. *Genome Medicine*, *9*, 109.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780.

Kazeminasab, S., Najmabadi, H., & Kahrizi, K. (2018). Intellectual disability and ataxia: Genetic collisions. *Archives of Iranian Medicine*, *21*, 29–40.

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*, 310–315.

Kobayashi, Y., Yang, S., Nykamp, K., Garcia, J., Lincoln, S. E., & Topper, S. E. (2017). Pathogenic variant burden in the ExAC database: An empirical approach to evaluating population data for clinical variant interpretation. *Genome Medicine*, *9*, 13.

Krumm, N., O'Roak, B. J., Shendure, J., & Eichler, E. E. (2014). A de novo convergence of autism genetics and molecular neuroscience. *Trends in Neurosciences*, *37*, 95–105.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., ... Monteiro, C. D. (2016). Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, *44*, W90–W97.

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., ... Birnbaum, D. P. (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., ... Stanley, F. (2011). Autism and intellectual disability are differentially related to sociodemographic background at birth. *PLOS One*, *6*, e17875.

Li, Q., & Wang, K. (2017). InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. *American Journal of Human Genetics*, *100*, 267–280.

Li, Y., Jia, X., Wu, H., Xun, G., Ou, J., Zhang, Q., ... Guo, H. (2018). Genotype and phenotype correlations for SHANK3 de novo mutations in neurodevelopmental disorders. *American Journal of Medical Genetics. Part A*, *176*, 2668–2676.

de Ligt, J., Willemsen, M. H., Bon, B. W. M., van, Kleefstra, T., Yntema, H. G., Kroes, T., ... Hoischen, A. (2012). Diagnostic exome sequencing in persons with severe intellectual disability. *New England Journal of Medicine*, *367*, 1921–1929.

Mata, I. F., Jang, Y., Kim, C.-H., Hanna, D. S., Dorschner, M. O., Samii, A., ... Factor, S. A. (2015). The RAB39B p.G192R mutation causes X-linked dominant Parkinson's disease. *Molecular Neurodegeneration*, *10*, 50.

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., ... Haft, D. R. (2018). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *46*, 726.

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. *Human Genome Variation*, *3*, 16029.

Neale, B. M., Kou, Y., Liu, L., Ma'ayan, A., Samocha, K. E., Sabo, A., ... Yoon, S. (2012). Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature*, *485*, 242–245.

O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L., Schwartz, J. J., Girirajan, S., ... Nickerson, D. A. (2011). Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics*, *43*, 585–589.

O'Roak, B. J., Vives, L., Fu, W., Egertson, J. D., Stanaway, I. B., Phelps, I. G., ... Hiatt, J. B. (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. *Nature Reviews Neuroscience*, 14, 383–400.

Pengelly, R. J., Greville-Heygate, S., Schmidt, S., Seaby, E. G., Jabalameli, M. R., Mehta, S. G., … Debant, A. (2016). Mutations specific to the Rac-GEF domain of TRIO cause intellectual disability and microcephaly. *Journal of Medical Genetics*, 53, 735–742.

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera – a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25, 1605–1612.

Pinto, D., Delaby, E., Merico, D., Barbosa, M., Merikangas, A., Klei, L., … Thompson, A. (2014). Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *American Journal of Human Genetics*, 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. *American Journal of Human Genetics*, 93, 368–383.

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, 20, 110–121.

Potenza, E., Di Domenico, T., Walsh, I., & Tosatto, S. C. E. (2015). MobiDB 2.0: An improved database of intrinsically disordered and mobile proteins. *Nucleic Acids Research*, 43, D315–D320.

Redin, C., Gérard, B., Lauer, J., Herenger, Y., Muller, J., Quartier, A., … Vicaire, S. (2014). Efficient strategy for the molecular diagnosis of intellectual disability using targeted high-throughput sequencing. *Journal of Medical Genetics*, 51, 724–736.

Reijnders, M. R. F., Kousi, M., Woerden, G. M., van, Klein, M., Bralten, J., Mancini, G. M. S., … Stevens, S. J. C. (2017). Variation in a range of mTOR-related genes associates with intracranial volume and intellectual disability. *Nature Communications*, 8, 1052.

Reva, B., Antipin, Y., & Sander, C. (2011). Predicting the functional impact of protein mutations: Application to cancer genomics. *Nucleic Acids Research*, 39, e118.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., … Rehm, H. L. (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. *Genetics in Medicine*, 17, 405–424.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29, 24–26.

Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., … Goodsell, D. S. (2017). The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Research*, 45, D271–D281.

Sadybekov, A., Tian, C., Arnesano, C., Katritch, V., & Herring, B. E. (2017). An autism spectrum disorder-related de novo mutation hotspot discovered in the GEF1 domain of Trio. *Nature Communications*, 8, 601.

Schneider, A., Dessimoz, C., & Gonnet, G. H. (2007). OMA Browser–exploring orthologous relations across 352 complete genomes. *Bioinformatics*, 23, 2180–2182.

Schwarz, J. M., Cooper, D. N., Schuelke, M., & Seelow, D. (2014). MutationTaster2: Mutation prediction for the deep-sequencing age. *Nature Methods*, 11, 361–362.

Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., … Gaunt, T. R. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Human Mutation*, 34, 57–65.

Shihab, H. A., Gough, J., Mort, M., Cooper, D. N., Day, I. N. M., & Gaunt, T. R. (2014). Ranking non-synonymous single nucleotide polymorphisms based on disease concepts. *Human Genomics*, 8, 11.

Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40, W452–W457.

Sirmaci, A., Spiliopoulos, M., Brancati, F., Powell, E., Duman, D., Abrams, A., … Blanton, S. (2011). Mutations in ANKRD11 cause KBG syndrome, characterized by intellectual disability, skeletal malformations, and macrodontia. *American Journal of Human Genetics*, 89, 289–294.

Swanger, S. A., Chen, W., Wells, G., Burger, P. B., Tankovic, A., Bhattacharya, S., … Millichap, J. J. (2016). Mechanistic insight into NMDA receptor dysregulation by rare variants in the GluN2A and GluN2B agonist binding domains. *American Journal of Human Genetics*, 99, 1261–1280.

Totsika, V., Hastings, R. P., Emerson, E., Lancaster, G. A., & Berridge, D. M. (2011). A population-based investigation of behavioural and emotional problems and maternal mental health: Associations with autism spectrum disorder and intellectual disability. *Journal of Child Psychology and Psychiatry and Allied Disciplines (Oxford)*, 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 Research*, 44, W117–W121.

Vissers, L. E. L. M., de Ligt, J., Gilissen, C., Janssen, I., Steehouwer, M., de Vries, P., … Hoischen, A. (2010). A de novo paradigm for mental retardation. *Nature Genetics*, 42, 1109–1112.

Woodbury-Smith, M., & Scherer, S. W. (2018). Progress in the genetics of autism spectrum disorder. *Developmental Medicine and Child Neurology*, 60, 445–451.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.