Copy number variants analysis in a cohort of isolated and syndromic developmental delay/intellectual disability reveals novel genomic disorders, position effects and candidate disease genes

CNVs analysis in a cohort of isolated and syndromic DD/ID reveals novel genomic disorders, position effects and candidate disease genes.

Eleonora Di Gregorio[1,2], Evelise Riberi[3], Elga Fabia Belligni[3], Elisa Biamino[3], Malte Spielmann[4], Ugo Ala[5,6], Alessandro Calcia[1], Irene Bagnasco[7], Diana Carli[1], Giorgia Gai[2], Mara Giordano[8], Andrea Guala[9], Roberto Keller[10], Giorgia Mandrile[2,11], Carlo Arduino[2], Antonella Maffè[12], Valeria Giorgia Naretto[2], Fabio Sirchia[1,2], Lorena Sorasio[13], Silvana Ungari[12], Andrea Zonta[2], Giulia Zacchetti[2,8], Flavia Talarico[2], Patrizia Pappi[2], Simona Cavalieri[1], Elisa Giorgio[1], Cecilia Mancini[1], Marta Ferrero[1], Alessandro Brussino[1], Elisa Savin[2], Marina Gandione[14], Alessandra Pelle[11,15], Daniela Francesca Giachino[11,15], Mario De Marchi[11,15], Gabriella Restagno[16], Paolo Provero[5,6], Margherita Cirillo Silengo[3], Enrico Grosso[2], Joseph D. Buxbaum[17,18,19,20,21,22], Barbara Pasini[1,2] , Silvia De Rubeis[17,18], Alfredo Brusco[1,2*] and Giovanni Battista Ferrero[3*]

*These authors equally contributed to the work.

[1] University of Torino, Department of Medical Sciences, 10126, Turin, Italy
[2] Città della Salute e della Scienza University Hospital, Medical Genetics Unit, 10126, Turin, Italy
[3] University of Torino, Department of Public Health and Pediatrics, 10126, Turin, Italy
[4] Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany
[5] Molecular Biotechnology Center (MBC), Computational Biology Unit, 10126, Turin, Italy
[6] University of Torino, Department of Molecular Biotechnology and Health Sciences, 10126, Turin, Italy
[7] Martini Hospital, ASL TO1, Neuropsichiatria Infantile, 10128, Turin, Italy
[8] University of Eastern Piedmont and Interdisciplinary Research Center of Autoimmune Diseases Department of Health Sciences, Laboratory of Genetics, 28100, Novara, Italy
[9] Castelli Hospital, SOC Pediatria, 28921, Verbania, Italy
[10] Adult Autism Center, Mental Health Department, ASL TO2, 10149, Turin, Italy
[11] San Luigi Gonzaga University Hospital, Medical Genetics, 10043, Orbassano (TO), Italy
[12] Santa Croce e Carle Hospital, Molecular Biology and Genetics Unit, 12100, Cuneo, Italy
[13] Santa Croce e Carle Hospital, Pediatrics, 12100, Cuneo, Italy
[14] University of Torino, Department of Neuropsychiatry, 10126, Turin, Italy
[15]University of Torino, Department of Clinical and Biological Sciences, 10100, Turin, Italy
[16]Città della Salute e della Scienza University Hospital, Laboratory of Molecular Genetics, 10126, Turin, Italy
[17] Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, 10029-5674, New York, USA
[18] Icahn School of Medicine at Mount Sinai, Department of Psychiatry, 10029-5674, New York, USA
[19] Icahn School of Medicine at Mount Sinai, Department of Neuroscience, 10029-5674, New York, USA.
[20] Icahn School of Medicine at Mount Sinai, Friedman Brain Institute, 10029-5674, New York, USA.

[21] Icahn School of Medicine at Mount Sinai, Department of Genetics and Genomic Sciences, 10029-5674, New York, USA.
[22] Icahn School of Medicine at Mount Sinai, Mindich Child Health and Development Institute, 10029-5674, New York, USA.

Corresponding authors:

Alfredo Brusco, University of Torino, Department of Medical Sciences, via Santena 19, 10126, Torino, Italy. E-mail: alfredo.brusco@unito.it

Giovanni Battista Ferrero, University of Torino, Department of Pediatric and Public Health Sciences, p.zza Polonia, 4, 10126, Torino, Italy. E-mail: giovannibattista.ferrero@unito.it

CONFLICT OF INTEREST STATEMENT

None of the authors has any conflict of interest to disclose.

ACKNOWLEDGMENTS

This study makes use of data generated by the DECIPHER community. A full list of centers that have contributed to the generation of the data is available at http://decipher.sanger.ac.uk and by email to decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

We are grateful to the participating families. This work was supported by the following grants: MURST60% and Associazione E.E.Rulfo (to AB), Associazione “Enrico e Ilaria sono con noi” ONLUS (to GBF), Fondazione FORMA (to GBF), DFG (SP1532/2-1) to MS. SDR is a fellow of the Seaver Foundation.

ABSTRACT

Array-comparative genomic hybridization (array-CGH) is a widely used technique to detect Copy Number Variants (CNVs) associated with developmental delay/intellectual disability (DD/ID). We performed a comprehensive array-CGH investigation of 1,015 consecutive cases with DD/ID and combined literature mining, genetic evidence, evolutionary constraint scores, and functional information in order to assess the pathogenicity of the CNVs. We identified non-benign CNVs in 29% of patients. Amongst the pathogenic variants (11%), detected with a yield consistent with the literature, we found rare genomic disorders and CNVs spanning known disease genes. We further identified and discussed 51 cases with likely pathogenic CNVs spanning novel candidate genes, including genes encoding synaptic components and/or proteins involved in corticogenesis. Additionally, we identified two deletions spanning potential Topological Associated Domain (TAD) boundaries likely affecting the regulatory landscape.

In conclusion, we show how phenotypic and genetic analyses of array-CGH data allow unraveling complex cases, identifying rare disease genes, and revealing unexpected position effects.

Keywords: array-CGH, CNV, genomic disorders, intellectual disability, developmental delay, autism spectrum disorder

INTRODUCTION

Developmental delay/intellectual disability (DD/ID) is a common neurodevelopmental disorder affecting 1–3% of children (1) that can co-manifest with autism spectrum disorder (ASD). The genetics of DD/ID is complex and includes chromosomal abnormalities, copy number variations (CNVs) and mutations in single genes (2). Despite the high prevalence in the population and the advances in the identification of risk genes and *loci* for DD/ID (3), only a fraction of patients can be identified with a specific genomic or monogenic disorder through array-comparative genomic hybridization (array-CGH), targeted sequencing of a panel of known risk genes or clinical whole-exome sequencing.

Array-CGH is the first tier diagnostic test, with a 5 to 10-fold improved detection rate over conventional karyotyping and an average diagnostic yield between 15–20% in patients with DD/ID (1). The interpretation of pathogenicity of the variants detected by array-CGH is often complicated (4) and requires the integration of multiple bioinformatics resources and a synergistic effort of a team of pediatricians, clinical geneticists and molecular biologists.

Here, we present an array-CGH study of an Italian cohort of over 1,000 cases diagnosed with DD/ID. We identified known genomic syndromes, including rare cases, likely pathogenic variants in known but yet poorly characterized genes, and variants in novel candidate genes. Furthermore, our study combined several bioinformatics analyses to identify novel likely pathogenic CNVs.

PATIENTS AND METHODS.

In a retrospective review of data, we examined the results of diagnostic array-CGH test on 1,015 consecutive cases diagnosed with DD/ID in the Medical Genetics Unit at the “Città della Salute e della Scienza” University Hospital, Turin, Italy, over a 7-year period (2008-2014). Array-CGH was performed using a clinically validated Agilent 60K platform, human assembly Feb. 2009 (GRCh37/hg19).

For the interpretation of pathogenicity, we integrated several tools.

We reported details on methods in the Supporting information.

RESULTS

We analyzed 1,015 patients referred to our genetic unit for idiopathic DD/ID with array-CGH. In 293 cases (178 males and 115 females), we identified at least one variant not reported as a polymorphism in Database of Genomics Variants (DGV). We then assigned these CNVs to one of two categories, pathogenic variants or variants of unknown significance (VOUS), further dividing VOUS into three subgroups (5) (Fig. 1). We will briefly discuss our findings, and report exemplifying cases.

Pathogenic genomic rearrangements

Among pathogenic CNVs, we included large genomic rearrangements (Table S1, see Supporting information), microdeletions and microduplications in *loci* associated with genomic disorders (Table S2), and CNVs in known dosage-sensitive genes (Table S3).

Specifically, we identified 28 cases with large genomic rearrangements spanning a region of 3 Mb or larger (28/293; 9.5%) (Table S1). Five subjects had derivative chromosomes inherited from parental balanced translocations, with some of them involving almost identical regions and thus cytogenetically undetectable (6). We also identified 58 microdeletions and 16 microduplications (74/293; 25.2%) associated with 39 known genomic disorders (Table S2). Of the 44 cases where inheritance could be tested, 26 were *de novo* and 18 inherited (13 from an asymptomatic parent and five from parents with subclinical phenotype). Among these, we included two subjects we described with novel genomic disorders: the centromeric 3q29 deletion syndrome (7) and the chromosome 17p13.1 duplication syndrome (8).

We found nine CNVs spanning known dosage-sensitive genes whose haploinsufficiency is associated with DD/ID and/or ASD (9/293; 3.1%) (Table S3). Seven were *de novo*, whereas one was inherited from an unaffected mother [*NRXN1* gene deletion (case 300674)] and one case was with unknown inheritance.

Variants of unknown significance (VOUS); likely pathogenic

Amongst the VOUS, we searched for genes with evidence of association to neurodevelopmental and psychiatric disorders, based on the published literature and consolidated gene sets, including risk genes for ASD (9), autosomal dominant or X-linked genes associated to brain developmental disorders as defined in the Gene2Phenotype database, and genome-wide association study (GWAS)-associated *loci* for schizophrenia (10) (see Supporting information). These analyses led to the identification of 52 likely pathogenic CNVs in 51 patients (51/1,015; 5.0%).

To further understand the neurobiology associated to these genes, we overlaid information on evolutionary constraint scores and functional categories relevant to neurodevelopmental disorders for all the genes spanned by these 52 variants, as described in Supporting information. We found one CNVs spanning *ERBB2IP* and one *NAA15*, both heavily constrained and targeted by *CHD8* (Fig. 2, Fig. S1). A likely pathogenic deletion spanning *ERBB2IP* (MIM#606944) was transmitted by an affected mother to a male with severe DD/ID, abnormal behavior, electroencephalogram (EEG) with focal epileptiform discharges, seizures, sensorineural hearing impairment, worsening myopia, downslanting palpebral fissures and proportionate short stature (case 263833) (Fig. S1A). His mother had a borderline ID, depression, sensorineural hearing impairment and epilepsy in infancy (not confirmed by EEG analysis). A *de novo* duplication encompassing *NAA15* (MIM#608000) was detected in a subject with reduced growth and weight gain in the first years of age, macrocephaly, abnormalities of the ectodermal tissues compatible with mild ectodermal dysplasia, and central hypothyroidism. Magnetic resonance showed hypoplasia of adenohypophysis and neurohypophysis,

and pituitary stalk hypoplasia (case 300731) (Fig. S1B). Noteworthy, *NAA15* has also been associated with congenital heart disease (11) and has been found with *de novo* mutations in individuals diagnosed with DD/ID (3).

A 105-kb deletion spanning exons 9-10 of the α–isoform and exons 8–9 of the β-isoform of *NRXN3* (MIM#600567) was found in a 7-year old girl with DD associated with ASD (case 300039). The patient, first child of non-consanguineous parents, was able to sit independently at 5-6 months of age, and started to walk unsupported at 16 months. At 24 months of age, she presented speech delay, DD and poor social interactions; ASD was diagnosed at three years old [Autism Diagnostic Observation Schedule (ADOS): language and communication: 7, reciprocal social interaction: 14, play: 4, stereotyped behaviors and restricted interests: 4]. We excluded Rett and Rett-like syndromes by Sanger sequencing of *MECP2*, *CDKL5*, *FOXG1*, and *MEF2C*. The deletion was inherited from the mother and shared with a maternal uncle; both individuals were healthy, suggesting incomplete penetrance.

A *de novo* 2.4 Mb deletion including *SNX3* (MIM#605930) was identified in a subject (case 296462) with microcephaly (head circumference: 47 cm, <3° centile, 5.5 years of age), thin hair, hypertelorism, epicanthus, craniosynostosis with proptosis due to orbital hypoplasia, depressed nasal bridge, smooth philtrum, thin lips, micrognathia, hypotonia and ligamentous laxity, and moderate DD/ID with severe dyspraxia. *SNX3* was listed as possible DD gene (Gene2Phenotype) based on a translocation disrupting the gene in a patient with ID and microcephaly, microphthalmia, ectrodactyly, and prognathism (12). This subject harbors two *de novo* deletions (6q21 and 14q32.2) and a paternally inherited 6q22.31 duplication. It is likely that multiple genes and/or interactions between genes contribute to the proband's phenotype.

We found two overlapping deletions (case 266246 and 300526), both maternally transmitted, in *TOP3B* (MIM#603582), a gene previously reported with a duplication in a patient with mild ID and generalized overgrowth (13). Subject 266246 inherited the deletion from the unaffected mother. He

showed a moderate ID, micrognathia, cerebellar vermis hypoplasia, corpus callosum hyperplasia, hypertelorism, synophrys, tall stature (>97th centile at 6 years), triangular face, low-set ears, anteverted nares, digital clubbing of the first finger of hands and feet. Individual 300526 inherited the deletion from the mother, who manifested with a moderate learning disability. He had language delay, hyperactivity, impulsivity, clumsiness, and abnormal facial features. *MECP2* gene mutations were excluded. Noteworthy, four identical deletions (324416, 318024, 307779, 300526) and one overlapping deletion (case 250737) (two inherited and the rest of unknown transmission) are reported in DECIPHER in individuals with DD/ID and/or ASD. Remarkably, *TOP3B* lies in the distal 22q11.2 microdeletion region and the majority of these rearrangements are due to non-allelic homologous misalignments and unequal recombination mediated by low-copy repeat sequences (14). Incomplete penetrance observed in our subject and in these cases is compatible with the lack of intolerance to LoF in the gene (Fig. 2).

We identified two cases with deletions encompassing *DLG2* (MIM#603583), a gene reported with deletions in DD/ID (15), and bipolar disorder (16). The first case is a 31-year old man (case 314659), first child of non-consanguineous parents, with negative family history, carrying a 139-kb deletion inherited from the unaffected father and spanning exon 5 of isoform 1 of *DLG2* (NM_001142699). The subject was born at full-term after an uneventful pregnancy, with normal auxometric parameters. Neuropsychiatric evaluation led to a diagnosis of high-functioning autism (intelligence quotient, IQ 79). The second patient (case 300042) was a 12-year old girl, first child of non-consanguineous parents, found with a 373-kb deletion spanning exon 7 of isoform 1 transmitted by the unaffected mother. The subject manifested with mild intellectual disability (total IQ 55, verbal IQ 55, performance IQ 70) and microcephaly (48.5 cm; < 3rd centile, 12 years of age), deeply set eyes, myopia, synophrys, hearing abnormality (mixed bilateral hearing loss) and short digits.

We found a likely pathogenic, 222-kb deletion encompassing *ASTN2* (MIM#612856), a gene reported with deletions in ASD (17) in a 23-year old woman (case 299884) with history of DD, microcephaly, mild intellectual disability (IQ 69, Terman-Merrill Intelligence Scale evaluation), and

lower limbs dysmetria. Magnetic resonance did not show cerebral anomalies. The subject was the first child of consanguineous parents (second cousins), and she inherited the deletion from the mother and shared it with a younger sister. Although we could not obtain a neuropsychiatric evaluation of the mother, she reported a need for special education in school. The younger, 21-years old sister showed a neuropsychiatric phenotype similar to that of the proband, with moderate intellectual disability (IQ 48), language delay, accompanied by strabismus and nystagmus on the right eye, mild thoracic scoliosis and hypertrichosis. Magnetic resonance revealed a ventricular enlargement and increased thickness of the corpus callosum. Based on the parents' consanguinity we cannot exclude a recessive disease as also demonstrated in (18).

One case with a CNV inherited from an affected parent is a 360-kb long, paternally transmitted duplication spanning *DRD5* and two other genes in a case with DD/ID (case 299797). The rearrangement overlaps with a larger duplication described in six related individuals, four of which had schizophrenia or schizoaffective disorder (the remaining two subjects were under the average age of onset in this family) (19).

Variants of unknown clinical significance; uncertain and likely benign

The remaining 152 CNVs were identified in 131 individuals (131/1,015; 13%) and classified as "VOUS uncertain" (109 CNVs, 99 patients; Table S6) and "VOUS likely benign" (43 CNVs, 32 patients; Table S7). Among the latters, four were X-linked duplications spanning gene deserts detected in males.

GO analyses

To gain functional insights in the genes spanning CNVs classified as VOUS potentially pathogenic and uncertain, we tested enrichment in Gene Ontology (GO) terms. We found 48 statistically significant (empirical p Value ≤ 0.01) GO entries. We ranked these GO selecting those with ≥ 2

observations (n= 24). In Table S5, we reported enriched GO terms associated with genes classified as likely pathogenic. Genes associated with the enriched GO categories were mainly involved in the essential metabolic pathway related to energy production affecting mitochondria function, and endosomal system. Alteration of these pathways in brain has been demonstrated to be common in ID and ASD pathogenesis (20, 21). In fact, brain energy supply is essential for a correct development and function of neurons, due to their high metabolic demand (20, 21).

Analysis of long distance regulatory domains

Among the large genomic deletions, we found a case (patient 314103) with clinical features inconsistent with the phenotype expected for the haploinsufficiency of the genes deleted in the rearrangement. The proband was a two-year old girl, first child of non-consanguineous parents (Fig. 3), with moderate ID, abnormality of cerebral white matter, defect of visual evoked potentials, hearing impairment and preauricular pits. Through array-CGH, we identified a *de novo* 8.1-Mb deletion on chromosome 8q12.3q13.3 spanning ~30 genes (Fig. 3). The deletion includes *EYA1*, a gene whose haploinsufficiency causes, among other related disorders, the Branchiootic syndrome (MIM#602588), an autosomal dominant disorder characterized by hearing loss, preauricular pits and mild DD in some cases. Based on the clinical phenotype and the genetic findings, the subject was diagnosed with Branchiootic syndrome. The region deleted also includes *SULF1* (MIM#610012) and *SLCO5A1* (NM_030958) (Fig. 3). Deletions spanning these two genes have been reported in four unrelated individuals with mesomelia-synostoses syndrome (MIM#600383) (22), a disorder characterized by skeletal anomalies, including mesomelic limb shortening and acral synostoses. However, our patient has no skeletal abnormalities, as demonstrated by X-ray scans. Additionally, DECIPHER inspection revealed three additional carriers of a larger deletion involving both *SULF1* and *SLCO5A1*, and all individuals share clinical traits of Branchiootic syndrome and no apparent skeletal phenotype (Fig. 3). Interestingly, bioinformatics analyses using the 3D Genome Browser (http://promoter.bx.psu.edu/hi-c/) revealed a potential Topological Associated Domain (TAD)

boundary within *SULF1* (Fig. 3). This raises the possibility that the mesomelia-synostoses phenotype described by Isidor and colleagues might result from secondary effects of the ablation of this TAD boundary rather than from *SULF1* and *SLCO5A1* haploinsufficiency, a model compatible with the lack of skeletal defects in our case.

We extended the search of TAD boundary deletions on the VOUS group. We identified a 284 kb deletion in an intergenic region distal to *CTNND2* (MIM#604275), a gene found deleted in ID and/or ASD (23) and recently proposed as an ASD risk gene (24). The deletion is predicted to disrupt a TAD boundary located upstream *CTNND2* and was found in a 19-year old woman (case 300081) with anorexia nervosa, anxiety disorder, dysmorphisms, tapered and long fingers and scoliosis. She was the second daughter of non-consanguineous parents, and the deletion was inherited from the mother and shared with the older sister, both suffering from psychological disturbances: the mother had anxiety disorder, while the sister had anorexia (Fig. 4). We hypothesize that the ablation of the boundary consequent to the deletion might result in the action of unintended enhancer elements on *CTNND2*, leading to altered regulation of *CTNND2*.

DISCUSSION

Our work reports the outcome of an array-CGH screening on an Italian cohort of 1,015 subjects diagnosed with DD/ID. We found 11% pathogenic CNVs and 18% of Variants of Unknown Significance (VOUS) in line with literature data (1). We further studied this latter group to correlate the CNV gene content with phenotype. We used annotations in gene classes related to DD/ID and ASD pathophysiology, i.e., genes regulated by the hub ASD genes *FMR1* (25) and *CHD8* (26), GO categories, ExAC constraint scores, TAD domains alteration, and segregation analysis to identify 52 likely pathogenic CNVs (5%) in 51 cases, providing novel insights into the clinical developmental role of candidate disease genes for DD/ID and ASD disorders.

Amongst these genes, several are human orthologues of mouse synaptic components (27) (*ASTN2*, *DLG2*, *GRIA2*, *ERBB2IP*, *SNX3* and *NRXN3*) and human genes encoding postsynaptic density (PSD) proteins (*DLG2*, *GRIA2*, *ERBB2IP*, *SNX3* and *NRXN3*) (28). Tests for higher cognitive functions performed in humans and mice with an impaired expression of *DLG2* (encoding PSD-93/chapsyn-110) showed defects in complex learning, cognitive flexibility and attention (29). Knockout mice for Erbin (encoded by the mouse orthologue of *ERBB2IP*) show behavioral disturbances, including impairment in forebrain neural circuits and sensory gating and poor motor coordination (30). Synaptic transmission can also be modulated trans-synaptically, as in the case of the pre-synaptic adhesion molecule Neurexin-3 (encoded by the mouse orthologue of *NRXN3*), which can control the post-synaptic expression of the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) glutamate receptors in the hippocampus (31).

All of these proteins are essential during the formation and maturation of synapses [Erbin (32), PSD-93/chapsyn-110 (33), Neurexin-3 (34)]. This evidence accrues on a robust body of evidence implicating synaptic function in ID and ASD pathophysiology (35). Further, some of the candidate genes we identified in the CNVs regulate brain development beyond synaptogenesis. For example, Astrotactin 2 (encoded by the mouse orthologue of *ASTN2*) is a membrane protein that mediates the formation of neuronal-glial adhesions during the migration of cerebellar granule neurons (36). Erbin is required for re-myelination of regenerated axons after injury (37). Interestingly, *ASTN2* and *SNX3* are also included in the endosome GO category, enriched in our analysis, supporting a known role for endocytosis mechanisms in neuron development and function (21).

Some of these genes encode for mRNAs that interact with the chromatin remodeler *CHD8* (*ERBB2IP*, *SNX3*) or with the translational regulator fragile X mental retardation protein (FMRP) (*DLG2*, *NRXN3*), suggesting that the synaptic expression and/or localization of these key synaptic components might undergo coordinated regulation.

Gene dosage alterations represent the most likely mechanism for the variants we identified, with most cases explained by haploinsufficiency of the candidate genes. It is also worth noticing that the

rearrangements in *DLG2* and *NRXN3* are transmitted by unaffected parents, indicating a variable expressivity and/or incomplete penetrance. This is in line with reports of intragenic deletions in *NRXN3* in individuals with ASD transmitted by parents with subclinical autistic traits or no clinical manifestations (38).

We also described two cases implicating regions susceptible to epigenetic dysregulation rather than altered gene dosage. One of our patients without skeletal anomalies had a *de novo* deletion overlapping the mesomelia-synostoses syndrome critical region (22). Since three cases with similar deletions in DECIPHER do not show sign of skeletal involvement, we formulate a hypothesis alternative to incomplete penetrance. We hypothesize that the deletions described by Isidor and colleagues disrupt a TAD boundary in one of the two genes (*SULF1*) they have proposed as haploinsufficient (Fig 3). We suggest that the ablation of this insulator element removes the constraints of a bone enhancer (likely acting on *SULF1* itself), allowing its action on genes located distally to *SULF1*. Based on the deletions reported in mesomelia-synostoses syndrome, a candidate gene is *NCOA2*, which encodes a transcriptional coactivator for nuclear hormone receptors (steroid, thyroid, retinoid, and vitamin D receptors) also known as TIF2. HEY1-NCOA2 fusion proteins have been found in specimens of mesenchymal chondrosarcoma, an aggressive, uncommon histologic tumor arising in bone and soft tissues (39). A similar mechanism has been proposed for another skeletal disorder, the mesomelic dysplasia Savarirayan type (MIM#645274). Three unrelated cases with 2-Mb overlapping *de novo* 6p22.3 microdeletions spanning four genes with no previous association with skeletal anomalies were identified (40). The discovery of a fourth individual with a larger deletion but no skeletal phenotype, as well as the identification of two TAD boundaries in the interval, suggest that the deletion might cause the aberrant activation of potential limb enhancers in the region (40).

In a second case, we found the deletion of a TAD boundary in proximity of *CTNND2* (Fig. 4). We propose that the deletion may cause altered expression of the encoded protein, δ-catenin, a regulator of synaptogenesis and dendritic morphogenesis cooperating with Erbin (32, 41). In fact, *Ctnnd2*

knockout mice exhibit structural and functional abnormalities of the synapses, as well as defects in spatial learning and fear conditioning (42). CNVs involving *CTNND2* have been described in individuals with ASD (24) and the gene has been proposed as an ASD risk gene based on rare mutations found in families with two or more severely affected females (24). Interestingly, the three cases in this family are female patients with psychological disturbances ranging from anorexia nervosa to anxiety disorder. This observation raises the intriguing possibility that loss-of-function or altered dosage of *CTNND2* might underlie DD/ID and/or ASD, while altered expression might result in a different phenotype.

Based on segregation analysis of a patient (case 296491) with an intragenic deletion of the *MBD5* gene, we classified deletions of exons 3-4 of the 5'UTR of this gene as likely benign. The deletion of exon 3 was shared with the father and paternal uncle, both unaffected (Fig. S2A). Further, DGV contains several unaffected controls carrying deletions in exons 3-4 (Fig. S2B). *MBD5* is considered a known disease gene (MIM#611472) and we think that reconsidering published cases with deletions in the 5'UTR may help to in further refine a consensus phenotype and explain cases of incomplete penetrance (43).

We also found six complex phenotypes with CNVs spanning a known disease gene or genomic disorder locus. These CNVs explain the clinical features only partially (Table S8). We think that these cases might be explained by compounding effects of the genomic disorder identified here and an independent Mendelian disorder, yet to be identified. Multiple molecular diagnoses have already been described in large surveys of patients analysed by diagnostic whole-exome sequencing, where they represent approximately 5% of solved cases (44).

In conclusion, clinical and genetic analysis of our cohort allowed identifying complex cases, possibly rare disease genes involved in DD/ID, and novel position effects.

FIGURE LEGENDS

Figure 1. Rationale for CNV interpretation.

The diagram summarizes the workflow and the findings of this study: 1,015 consecutive ID/DD cases were studied by 60K array-CGH. CNVs included in DGV as polymorphisms were considered benign, and the remaining CNVs were initially divided into pathogenic and VOUS, based on MIM# and literature data. Pathogenic CNVs were assigned to this category because considered large (> 3 Mb) and containing gene(s) relevant for the pathology, spanning known genomic disorders regions or known dosage sensitive disease genes. VOUS were divided into three categories exploiting several bioinformatics tools among which: Database of Genomics Variants (http://dgv.tcag.ca/dgv/app/home), DECIPHER, OMIM: Online Mendelian Inheritance in Man (www.omim.org). SFARI, NDAR, NDD genes and loci, ExAC, GO.

Figure 2. Evolutionary constraints and functional annotation of genes in likely pathogenic CNVs.

The scatter plot shows the UCSC genes spanned by the CNVs distributed based on their ExAC pLI (y-axis) and missense z-score (x-axis). Dotted lines represent thresholds for significant constraints. Relevant genes discussed in the text are indicated. Genes were also annotated based on the following functional categories: ASD risk genes (blue), DD risk genes (light blue), GWAS-associated schizophrenia *loci* (green), genes encoding synaptic proteins (orange), genes encoding PSD proteins (purple), genes encoding FMRP targets (gold), and genes encoding *CHD8* targets (beige).

Figure 3. TAD and chromosome 8q12.3q13.3 deletion.

Pedigree of case 314103 (arrow). Filled symbols: affected subjects; empty symbols: unaffected subjects; del: deletion; dup: duplication. The candidate gene is indicated in bold; blue bar: duplication;

red bar: deletion. On the right, deleted region in case 314103 and Hi-C profile derived from human ES cells associated to this region. Hi-C interactions are shown in a heatmap with each dot reflecting two interaction pairs. The resulting interaction profile shows the formation of triangles (schematically enhanced in color) that represent individual TADs. Dashed lines indicate predicted TAD boundaries. Red, deletion in our subject; grey, cases with an overlapping deletion reported in DECIPHER and published literature (22). The genes spanned in the 8q12.3q13.3 region between 62.5 and 72.5 Mb (hg19) are shown below. *SULF1* and *SLCO5A1*, genes associated with Mesomelia synostoses syndrome, and *EYA1* causing Branchiootic syndrome, are in bold. Below, H3K4Me1 and H3K27Ac chromatin signatures (indicative of active/poised enhancers) in the genomic region between 69.5 and 69.85 Mb.

Figure 4. TAD and chromosome 5p15.2 deletion.

Pedigree of case 300081 (see previous legend). On the right, deleted region in case 300081 and Hi-C profile derived from human ES cells, as described above. Dashed lines indicate predicted TAD boundaries. The red bar indicates the deletion in our subject. Genes located in 5p15.2 genomic region between 11.5 and 13.5 Mb (hg19) are shown, with *CTNND2,* associated with psychiatric disturbances, in bold. H3K4Me1 and H3K27Ac chromatin signatures in the genomic region between 11.5 and 13.5 Mb at chromosome 5p are shown.

REFERENCES

1. Miller DT, Adam MP, Aradhya S et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 2010: 86: 749-764.
2. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annual review of genetics 2011: 45: 81-104.
3. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature 2015: 519: 223-228.
4. South ST, Lee C, Lamb AN et al. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genetics in medicine : official journal of the American College of Medical Genetics 2013: 15: 901-909.
5. Kearney HM, South ST, Wolff DJ et al. American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. Genetics in medicine : official journal of the American College of Medical Genetics 2011: 13: 676-679.
6. Di Gregorio E, Savin E, Biamino E et al. Large cryptic genomic rearrangements with apparently normal karyotypes detected by array-CGH. Mol Cytogenet 2014: 7: 82.
7. Biamino E, Di Gregorio E, Belligni EF et al. A novel 3q29 deletion associated with autism, intellectual disability, psychiatric disorders, and obesity. Am J Med Genet B Neuropsychiatr Genet 2016: 171: 290-299.
8. Belligni EF, Di Gregorio E, Biamino E et al. 790 Kb microduplication in chromosome band 17p13.1 associated with intellectual disability, afebrile seizures, dysmorphic features, diabetes, and hypothyroidism. Eur J Med Genet 2012: 55: 222-224.
9. Sanders SJ, He X, Willsey AJ et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 2015: 87: 1215-1233.
10. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014: 511: 421-427.
11. Zaidi S, Choi M, Wakimoto H et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature 2013: 498: 220-223.
12. Vervoort VS, Viljoen D, Smart R et al. Sorting nexin 3 (SNX3) is disrupted in a patient with a translocation t(6;13) (q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype. Journal of medical genetics 2002: 39: 893-899.
13. Tarsitano M, Ceglia C, Novelli A et al. Microduplications in 22q11.2 and 8q22.1 associated with mild mental retardation and generalized overgrowth. Gene 2014: 536: 213-216.
14. Shaikh TH, O'Connor RJ, Pierpont ME et al. Low copy repeats mediate distal chromosome 22q11.2 deletions: sequence analysis predicts breakpoint mechanisms. Genome Res 2007: 17: 482-491.
15. Sahoo T, Theisen A, Rosenfeld JA et al. Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. Genetics in medicine : official journal of the American College of Medical Genetics 2011: 13: 868-880.
16. Noor A, Lionel AC, Cohen-Woods S et al. Copy number variant study of bipolar disorder in Canadian and UK populations implicates synaptic genes. Am J Med Genet B Neuropsychiatr Genet 2014: 165B: 303-313.
17. Glessner JT, Wang K, Cai G et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 2009: 459: 569-573.

18. Giorgio E, Ciolfi A, Biamino E et al. Whole exome sequencing is necessary to clarify ID/DD cases with de novo copy number variants of uncertain significance: Two proof-of-concept examples. Am J Med Genet A 2016: 170: 1772-1779.
19. Van Den Bossche MJ, Strazisar M, Cammaerts S et al. Identification of rare copy number variants in high burden schizophrenia families. Am J Med Genet B Neuropsychiatr Genet 2013: 162B: 273-282.
20. Valenti D, de Bari L, de Filippis B et al. Mitochondrial dysfunction as a central actor in intellectual disability-related diseases: An overview of Down syndrome, autism, Fragile X and Rett syndrome. Neurosci Biobehav R 2014: 46: 202-217.
21. Patak J, Zhang-James Y, Faraone SV. Endosomal system genetics and autism spectrum disorders: A literature review. Neurosci Biobehav R 2016: 65: 95-112.
22. Isidor B, Pichon O, Redon R et al. Mesomelia-synostoses syndrome results from deletion of SULF1 and SLCO5A1 genes at 8q13. Am J Hum Genet 2010: 87: 95-100.
23. Hofmeister W, Nilsson D, Topa A et al. CTNND2-a candidate gene for reading problems and mild intellectual disability. J Med Genet 2015: 52: 111-122.
24. Turner TN, Sharma K, Oh EC et al. Loss of delta-catenin function in severe autism. Nature 2015: 520: 51-56.
25. Darnell JC, Van Driesche SJ, Zhang C et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 2011: 146: 247-261.
26. Cotney J, Muhle RA, Sanders SJ et al. The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat Commun 2015: 6: 6404.
27. Collins MO, Husi H, Yu L et al. Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. J Neurochem 2006: 97 Suppl 1: 16-23.
28. Bayes A, van de Lagemaat LN, Collins MO et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci 2011: 14: 19-21.
29. Nithianantharajah J, Komiyama NH, McKechanie A et al. Synaptic scaffold evolution generated components of vertebrate cognitive complexity. Nat Neurosci 2013: 16: 16-24.
30. Tao Y, Chen YJ, Shen C et al. Erbin interacts with TARP gamma-2 for surface expression of AMPA receptors in cortical interneurons. Nat Neurosci 2013: 16: 290-299.
31. Aoto J, Foldy C, Ilcus SM et al. Distinct circuit-dependent functions of presynaptic neurexin-3 at GABAergic and glutamatergic synapses. Nat Neurosci 2015: 18: 997-1007.
32. Arikkath J, Israely I, Tao Y et al. Erbin controls dendritic morphogenesis by regulating localization of delta-catenin. J Neurosci 2008: 28: 7047-7056.
33. Elias GM, Funke L, Stein V et al. Synapse-specific and developmentally regulated targeting of AMPA receptors by a family of MAGUK scaffolding proteins. Neuron 2006: 52: 307-320.
34. Dean C, Scholl FG, Choih J et al. Neurexin mediates the assembly of presynaptic terminals. Nat Neurosci 2003: 6: 708-716.
35. De Rubeis S, He X, Goldberg AP et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014: 515: 209-215.
36. Wilson PM, Fryer RH, Fang Y et al. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. J Neurosci 2010: 30: 8529-8540.
37. Liang C, Tao Y, Shen C et al. Erbin is required for myelination in regenerated axons after injury. J Neurosci 2012: 32: 15169-15180.
38. Vaags AK, Lionel AC, Sato D et al. Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am J Hum Genet 2012: 90: 133-141.
39. Bishop MW, Somerville JM, Bahrami A et al. Mesenchymal Chondrosarcoma in Children and Young Adults: A Single Institution Retrospective Review. Sarcoma 2015: 2015: 608279.
40. Flottmann R, Wagner J, Kobus K et al. Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. J Med Genet 2015: 52: 476-483.

41. Arikkath J, Peng IF, Ng YG et al. Delta-catenin regulates spine and synapse morphogenesis and function in hippocampal neurons during development. J Neurosci 2009: 29: 5435-5442.
42. Matter C, Pribadi M, Liu X et al. Delta-catenin is required for the maintenance of neural structure and function in mature cortex in vivo. Neuron 2009: 64: 320-327.
43. Hodge JC, Mitchell E, Pillalamarri V et al. Disruption of MBD5 contributes to a spectrum of psychopathology and neurodevelopmental abnormalities. Mol Psychiatry 2014: 19: 368-379.
44. Posey JE, Harel T, Liu P et al. Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. The New England journal of medicine 2017: 376: 21-31.

CNVs analysis in a cohort of isolated and syndromic DD/ID reveals novel genomic disorders, position effects and candidate disease genes.

Eleonora Di Gregorio[1,2], Evelise Riberi[3], Elga Fabia Belligni[3], Elisa Biamino[3], Malte Spielmann[4], Ugo Ala[5,6], Alessandro Calcia[1], Irene Bagnasco[7], Diana Carli[1], Giorgia Gai[2], Mara Giordano[8], Andrea Guala[9], Roberto Keller[10], Giorgia Mandrile[2,11], Carlo Arduino[2], Antonella Maffè[12], Valeria Giorgia Naretto[2], Fabio Sirchia[1,2], Lorena Sorasio[13], Silvana Ungari[12], Andrea Zonta[2], Giulia Zacchetti[2,8], Flavia Talarico[2], Patrizia Pappi[2], Simona Cavalieri[1], Elisa Giorgio[1], Cecilia Mancini[1], Marta Ferrero[1], Alessandro Brussino[1], Elisa Savin[2], Marina Gandione[14], Alessandra Pelle[11,15], Daniela Francesca Giachino[11,15], Mario De Marchi[11,15], Gabriella Restagno[16], Paolo Provero[5,6], Margherita Cirillo Silengo[3], Enrico Grosso[2], Joseph D. Buxbaum[17,18,19,20,21,22], Barbara Pasini[1,2] , Silvia De Rubeis[17,18], Alfredo Brusco[1,2*] and Giovanni Battista Ferrero[3*]

*These authors equally contributed to the work.

[1] University of Torino, Department of Medical Sciences, 10126, Turin, Italy
[2] Città della Salute e della Scienza University Hospital, Medical Genetics Unit, 10126, Turin, Italy
[3] University of Torino, Department of Public Health and Pediatrics, 10126, Turin, Italy
[4] Max Planck Institute for Molecular Genetics, 14195, Berlin, Germany
[5] Molecular Biotechnology Center (MBC), Computational Biology Unit, 10126, Turin, Italy
[6] University of Torino, Department of Molecular Biotechnology and Health Sciences, 10126, Turin, Italy
[7] Martini Hospital, ASL TO1, Neuropsichiatria Infantile, 10128, Turin, Italy
[8] University of Eastern Piedmont and Interdisciplinary Research Center of Autoimmune Diseases Department of Health Sciences, Laboratory of Genetics, 28100, Novara, Italy
[9] Castelli Hospital, SOC Pediatria, 28921, Verbania, Italy
[10] Adult Autism Center, Mental Health Department, ASL TO2, 10149, Turin, Italy
[11] San Luigi Gonzaga University Hospital, Medical Genetics, 10043, Orbassano (TO), Italy
[12] Santa Croce e Carle Hospital, Molecular Biology and Genetics Unit, 12100, Cuneo, Italy
[13] Santa Croce e Carle Hospital, Pediatrics, 12100, Cuneo, Italy
[14] University of Torino, Department of Neuropsychiatry, 10126, Turin, Italy

[15]University of Torino, Department of Clinical and Biological Sciences, 10100, Turin, Italy
[16]Città della Salute e della Scienza University Hospital, Laboratory of Molecular Genetics, 10126, Turin, Italy
[17] Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, 10029-5674, New York, USA
[18] Icahn School of Medicine at Mount Sinai, Department of Psychiatry, 10029-5674, New York, USA
[19] Icahn School of Medicine at Mount Sinai, Department of Neuroscience, 10029-5674, New York, USA.
[20] Icahn School of Medicine at Mount Sinai, Friedman Brain Institute, 10029-5674, New York, USA.
[21] Icahn School of Medicine at Mount Sinai, Department of Genetics and Genomic Sciences, 10029-5674, New York, USA.
[22] Icahn School of Medicine at Mount Sinai, Mindich Child Health and Development Institute, 10029-5674, New York, USA.

Corresponding authors:

Alfredo Brusco, University of Torino, Department of Medical Sciences, via Santena 19, 10126, Torino, Italy. E-mail: alfredo.brusco@unito.it

Giovanni Battista Ferrero, University of Torino, Department of Pediatric and Public Health Sciences, p.zza Polonia, 4, 10126, Torino, Italy. E-mail: giovannibattista.ferrero@unito.it

CONFLICT OF INTEREST STATEMENT

None of the authors has any conflict of interest to disclose.

ACKNOWLEDGMENTS

This study makes use of data generated by the DECIPHER community. A full list of centers that have contributed to the generation of the data is available at http://decipher.sanger.ac.uk and by email to decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

We are grateful to the participating families. This work was supported by the following grants: MURST60% and Associazione E.E.Rulfo (to AB), Associazione “Enrico e Ilaria sono con noi” ONLUS (to GBF), Fondazione FORMA (to GBF), DFG (SP1532/2-1) to MS. SDR is a fellow of the Seaver Foundation.

ABSTRACT

Array-comparative genomic hybridization (array-CGH) is a widely used technique to detect Copy Number Variants (CNVs) associated with developmental delay/intellectual disability (DD/ID). We performed a comprehensive array-CGH investigation of 1,015 consecutive cases with DD/ID and combined literature mining, genetic evidence, evolutionary constraint scores, and functional information in order to assess the pathogenicity of the CNVs. We identified non-benign CNVs in 29% of patients. Amongst the pathogenic variants (11%), detected with a yield consistent with the literature, we found rare genomic disorders and CNVs spanning known disease genes. We further identified and discussed 51 cases with likely pathogenic CNVs spanning novel candidate genes, including genes encoding synaptic components and/or proteins involved in corticogenesis. Additionally, we identified two deletions spanning potential Topological Associated Domain (TAD) boundaries likely affecting the regulatory landscape.

In conclusion, we show how phenotypic and genetic analyses of array-CGH data allow unraveling complex cases, identifying rare disease genes, and revealing unexpected position effects.

Keywords: array-CGH, CNV, genomic disorders, intellectual disability, developmental delay, autism spectrum disorder

INTRODUCTION

Developmental delay/intellectual disability (DD/ID) is a common neurodevelopmental disorder affecting 1–3% of children (1) that can co-manifest with autism spectrum disorder (ASD). The genetics of DD/ID is complex and includes chromosomal abnormalities, copy number variations (CNVs) and mutations in single genes (2). Despite the high prevalence in the population and the advances in the identification of risk genes and *loci* for DD/ID (3), only a fraction of patients can be identified with a specific genomic or monogenic disorder through array-comparative genomic hybridization (array-CGH), targeted sequencing of a panel of known risk genes or clinical whole-exome sequencing.

Array-CGH is the first tier diagnostic test, with a 5 to 10-fold improved detection rate over conventional karyotyping and an average diagnostic yield between 15–20% in patients with DD/ID (1). The interpretation of pathogenicity of the variants detected by array-CGH is often complicated (4) and requires the integration of multiple bioinformatics resources and a synergistic effort of a team of pediatricians, clinical geneticists and molecular biologists.

Here, we present an array-CGH study of an Italian cohort of over 1,000 cases diagnosed with DD/ID. We identified known genomic syndromes, including rare cases, likely pathogenic variants in known but yet poorly characterized genes, and variants in novel candidate genes. Furthermore, our study combined several bioinformatics analyses to identify novel likely pathogenic CNVs.

PATIENTS AND METHODS.

In a retrospective review of data, we examined the results of diagnostic array-CGH test on 1,015 consecutive cases diagnosed with DD/ID in the Medical Genetics Unit at the “Città della Salute e della Scienza” University Hospital, Turin, Italy, over a 7-year period (2008-2014). Array-CGH was performed using a clinically validated Agilent 60K platform, human assembly Feb. 2009 (GRCh37/hg19).

For the interpretation of pathogenicity, we integrated several tools.

We reported details on methods in the Supporting information.

RESULTS

We analyzed 1,015 patients referred to our genetic unit for idiopathic DD/ID with array-CGH. In 293 cases (178 males and 115 females), we identified at least one variant not reported as a polymorphism in Database of Genomics Variants (DGV). We then assigned these CNVs to one of two categories, pathogenic variants or variants of unknown significance (VOUS), further dividing VOUS into three subgroups (5) (Fig. 1). We will briefly discuss our findings, and report exemplifying cases.

Pathogenic genomic rearrangements

Among pathogenic CNVs, we included large genomic rearrangements (Table S1, see Supporting information), microdeletions and microduplications in *loci* associated with genomic disorders (Table S2), and CNVs in known dosage-sensitive genes (Table S3).

Specifically, we identified 28 cases with large genomic rearrangements spanning a region of 3 Mb or larger (28/293; 9.5%) (Table S1). Five subjects had derivative chromosomes inherited from parental balanced translocations, with some of them involving almost identical regions and thus cytogenetically undetectable (6). We also identified 58 microdeletions and 16 microduplications (74/293; 25.2%) associated with 39 known genomic disorders (Table S2). Of the 44 cases where inheritance could be tested, 26 were *de novo* and 18 inherited (13 from an asymptomatic parent and five from parents with subclinical phenotype). Among these, we included two subjects we described with novel genomic disorders: the centromeric 3q29 deletion syndrome (7) and the chromosome 17p13.1 duplication syndrome (8).

We found nine CNVs spanning known dosage-sensitive genes whose haploinsufficiency is associated with DD/ID and/or ASD (9/293; 3.1%) (Table S3). Seven were *de novo*, whereas one was inherited from an unaffected mother [*NRXN1* gene deletion (case 300674)] and one case was with unknown inheritance.

Variants of unknown significance (VOUS); likely pathogenic

Amongst the VOUS, we searched for genes with evidence of association to neurodevelopmental and psychiatric disorders, based on the published literature and consolidated gene sets, including risk genes for ASD (9), autosomal dominant or X-linked genes associated to brain developmental disorders as defined in the Gene2Phenotype database, and genome-wide association study (GWAS)-associated *loci* for schizophrenia (10) (see Supporting information). These analyses led to the identification of 52 likely pathogenic CNVs in 51 patients (51/1,015; 5.0%).

To further understand the neurobiology associated to these genes, we overlaid information on evolutionary constraint scores and functional categories relevant to neurodevelopmental disorders for all the genes spanned by these 52 variants, as described in Supporting information. We found one CNVs spanning *ERBB2IP* and one *NAA15*, both heavily constrained and targeted by *CHD8* (Fig. 2, Fig. S1). A likely pathogenic deletion spanning *ERBB2IP* (MIM#606944) was transmitted by an affected mother to a male with severe DD/ID, abnormal behavior, electroencephalogram (EEG) with focal epileptiform discharges, seizures, sensorineural hearing impairment, worsening myopia, downslanting palpebral fissures and proportionate short stature (case 263833) (Fig. S1A). His mother had a borderline ID, depression, sensorineural hearing impairment and epilepsy in infancy (not confirmed by EEG analysis). A *de novo* duplication encompassing *NAA15* (MIM#608000) was detected in a subject with reduced growth and weight gain in the first years of age, macrocephaly, abnormalities of the ectodermal tissues compatible with mild ectodermal dysplasia, and central hypothyroidism. Magnetic resonance showed hypoplasia of adenohypophysis and neurohypophysis,

and pituitary stalk hypoplasia (case 300731) (Fig. S1B). Noteworthy, *NAA15* has also been associated with congenital heart disease (11) and has been found with *de novo* mutations in individuals diagnosed with DD/ID (3).

A 105-kb deletion spanning exons 9-10 of the α–isoform and exons 8–9 of the β-isoform of *NRXN3* (MIM#600567) was found in a 7-year old girl with DD associated with ASD (case 300039). The patient, first child of non-consanguineous parents, was able to sit independently at 5-6 months of age, and started to walk unsupported at 16 months. At 24 months of age, she presented speech delay, DD and poor social interactions; ASD was diagnosed at three years old [Autism Diagnostic Observation Schedule (ADOS): language and communication: 7, reciprocal social interaction: 14, play: 4, stereotyped behaviors and restricted interests: 4]. We excluded Rett and Rett-like syndromes by Sanger sequencing of *MECP2*, *CDKL5*, *FOXG1*, and *MEF2C*. The deletion was inherited from the mother and shared with a maternal uncle; both individuals were healthy, suggesting incomplete penetrance.

A *de novo* 2.4 Mb deletion including *SNX3* (MIM#605930) was identified in a subject (case 296462) with microcephaly (head circumference: 47 cm, <3° centile, 5.5 years of age), thin hair, hypertelorism, epicanthus, craniosynostosis with proptosis due to orbital hypoplasia, depressed nasal bridge, smooth philtrum, thin lips, micrognathia, hypotonia and ligamentous laxity, and moderate DD/ID with severe dyspraxia. *SNX3* was listed as possible DD gene (Gene2Phenotype) based on a translocation disrupting the gene in a patient with ID and microcephaly, microphthalmia, ectrodactyly, and prognathism (12). This subject harbors two *de novo* deletions (6q21 and 14q32.2) and a paternally inherited 6q22.31 duplication. It is likely that multiple genes and/or interactions between genes contribute to the proband's phenotype.

We found two overlapping deletions (case 266246 and 300526), both maternally transmitted, in *TOP3B* (MIM#603582), a gene previously reported with a duplication in a patient with mild ID and generalized overgrowth (13). Subject 266246 inherited the deletion from the unaffected mother. He

showed a moderate ID, micrognathia, cerebellar vermis hypoplasia, corpus callosum hyperplasia, hypertelorism, synophrys, tall stature (>97th centile at 6 years), triangular face, low-set ears, anteverted nares, digital clubbing of the first finger of hands and feet. Individual 300526 inherited the deletion from the mother, who manifested with a moderate learning disability. He had language delay, hyperactivity, impulsivity, clumsiness, and abnormal facial features. *MECP2* gene mutations were excluded. Noteworthy, four identical deletions (324416, 318024, 307779, 300526) and one overlapping deletion (case 250737) (two inherited and the rest of unknown transmission) are reported in DECIPHER in individuals with DD/ID and/or ASD. Remarkably, *TOP3B* lies in the distal 22q11.2 microdeletion region and the majority of these rearrangements are due to non-allelic homologous misalignments and unequal recombination mediated by low-copy repeat sequences (14). Incomplete penetrance observed in our subject and in these cases is compatible with the lack of intolerance to LoF in the gene (Fig. 2).

We identified two cases with deletions encompassing *DLG2* (MIM#603583), a gene reported with deletions in DD/ID (15), and bipolar disorder (16). The first case is a 31-year old man (case 314659), first child of non-consanguineous parents, with negative family history, carrying a 139-kb deletion inherited from the unaffected father and spanning exon 5 of isoform 1 of *DLG2* (NM_001142699). The subject was born at full-term after an uneventful pregnancy, with normal auxometric parameters. Neuropsychiatric evaluation led to a diagnosis of high-functioning autism (intelligence quotient, IQ 79). The second patient (case 300042) was a 12-year old girl, first child of non-consanguineous parents, found with a 373-kb deletion spanning exon 7 of isoform 1 transmitted by the unaffected mother. The subject manifested with mild intellectual disability (total IQ 55, verbal IQ 55, performance IQ 70) and microcephaly (48.5 cm; < 3rd centile, 12 years of age), deeply set eyes, myopia, synophrys, hearing abnormality (mixed bilateral hearing loss) and short digits.

We found a likely pathogenic, 222-kb deletion encompassing *ASTN2* (MIM#612856), a gene reported with deletions in ASD (17) in a 23-year old woman (case 299884) with history of DD, microcephaly, mild intellectual disability (IQ 69, Terman-Merrill Intelligence Scale evaluation), and

lower limbs dysmetria. Magnetic resonance did not show cerebral anomalies. The subject was the first child of consanguineous parents (second cousins), and she inherited the deletion from the mother and shared it with a younger sister. Although we could not obtain a neuropsychiatric evaluation of the mother, she reported a need for special education in school. The younger, 21-years old sister showed a neuropsychiatric phenotype similar to that of the proband, with moderate intellectual disability (IQ 48), language delay, accompanied by strabismus and nystagmus on the right eye, mild thoracic scoliosis and hypertrichosis. Magnetic resonance revealed a ventricular enlargement and increased thickness of the corpus callosum. Based on the parents' consanguinity we cannot exclude a recessive disease as also demonstrated in (18).

One case with a CNV inherited from an affected parent is a 360-kb long, paternally transmitted duplication spanning *DRD5* and two other genes in a case with DD/ID (case 299797). The rearrangement overlaps with a larger duplication described in six related individuals, four of which had schizophrenia or schizoaffective disorder (the remaining two subjects were under the average age of onset in this family) (19).

Variants of unknown clinical significance; uncertain and likely benign

The remaining 152 CNVs were identified in 131 individuals (131/1,015; 13%) and classified as "VOUS uncertain" (109 CNVs, 99 patients; Table S6) and "VOUS likely benign" (43 CNVs, 32 patients; Table S7). Among the latters, four were X-linked duplications spanning gene deserts detected in males.

GO analyses

To gain functional insights in the genes spanning CNVs classified as VOUS potentially pathogenic and uncertain, we tested enrichment in Gene Ontology (GO) terms. We found 48 statistically significant (empirical p Value ≤ 0.01) GO entries. We ranked these GO selecting those with ≥ 2

observations (n= 24). In Table S5, we reported enriched GO terms associated with genes classified as likely pathogenic. Genes associated with the enriched GO categories were mainly involved in the essential metabolic pathway related to energy production affecting mitochondria function, and endosomal system. Alteration of these pathways in brain has been demonstrated to be common in ID and ASD pathogenesis (20, 21). In fact, brain energy supply is essential for a correct development and function of neurons, due to their high metabolic demand (20, 21).

Analysis of long distance regulatory domains

Among the large genomic deletions, we found a case (patient 314103) with clinical features inconsistent with the phenotype expected for the haploinsufficiency of the genes deleted in the rearrangement. The proband was a two-year old girl, first child of non-consanguineous parents (Fig. 3), with moderate ID, abnormality of cerebral white matter, defect of visual evoked potentials, hearing impairment and preauricular pits. Through array-CGH, we identified a *de novo* 8.1-Mb deletion on chromosome 8q12.3q13.3 spanning ~30 genes (Fig. 3). The deletion includes *EYA1*, a gene whose haploinsufficiency causes, among other related disorders, the Branchiootic syndrome (MIM#602588), an autosomal dominant disorder characterized by hearing loss, preauricular pits and mild DD in some cases. Based on the clinical phenotype and the genetic findings, the subject was diagnosed with Branchiootic syndrome. The region deleted also includes *SULF1* (MIM#610012) and *SLCO5A1* (NM_030958) (Fig. 3). Deletions spanning these two genes have been reported in four unrelated individuals with mesomelia-synostoses syndrome (MIM#600383) (22), a disorder characterized by skeletal anomalies, including mesomelic limb shortening and acral synostoses. However, our patient has no skeletal abnormalities, as demonstrated by X-ray scans. Additionally, DECIPHER inspection revealed three additional carriers of a larger deletion involving both *SULF1* and *SLCO5A1*, and all individuals share clinical traits of Branchiootic syndrome and no apparent skeletal phenotype (Fig. 3). Interestingly, bioinformatics analyses using the 3D Genome Browser (http://promoter.bx.psu.edu/hi-c/) revealed a potential Topological Associated Domain (TAD)

boundary within *SULF1* (Fig. 3). This raises the possibility that the mesomelia-synostoses phenotype described by Isidor and colleagues might result from secondary effects of the ablation of this TAD boundary rather than from *SULF1* and *SLCO5A1* haploinsufficiency, a model compatible with the lack of skeletal defects in our case.

We extended the search of TAD boundary deletions on the VOUS group. We identified a 284 kb deletion in an intergenic region distal to *CTNND2* (MIM#604275), a gene found deleted in ID and/or ASD (23) and recently proposed as an ASD risk gene (24). The deletion is predicted to disrupt a TAD boundary located upstream *CTNND2* and was found in a 19-year old woman (case 300081) with anorexia nervosa, anxiety disorder, dysmorphisms, tapered and long fingers and scoliosis. She was the second daughter of non-consanguineous parents, and the deletion was inherited from the mother and shared with the older sister, both suffering from psychological disturbances: the mother had anxiety disorder, while the sister had anorexia (Fig. 4). We hypothesize that the ablation of the boundary consequent to the deletion might result in the action of unintended enhancer elements on *CTNND2*, leading to altered regulation of *CTNND2*.

DISCUSSION

Our work reports the outcome of an array-CGH screening on an Italian cohort of 1,015 subjects diagnosed with DD/ID. We found 11% pathogenic CNVs and 18% of Variants of Unknown Significance (VOUS) in line with literature data (1). We further studied this latter group to correlate the CNV gene content with phenotype. We used annotations in gene classes related to DD/ID and ASD pathophysiology, i.e., genes regulated by the hub ASD genes *FMR1* (25) and *CHD8* (26), GO categories, ExAC constraint scores, TAD domains alteration, and segregation analysis to identify 52 likely pathogenic CNVs (5%) in 51 cases, providing novel insights into the clinical developmental role of candidate disease genes for DD/ID and ASD disorders.

Amongst these genes, several are human orthologues of mouse synaptic components (27) (*ASTN2*, *DLG2*, *GRIA2*, *ERBB2IP*, *SNX3* and *NRXN3*) and human genes encoding postsynaptic density (PSD) proteins (*DLG2*, *GRIA2*, *ERBB2IP*, *SNX3* and *NRXN3*) (28). Tests for higher cognitive functions performed in humans and mice with an impaired expression of *DLG2* (encoding PSD-93/chapsyn-110) showed defects in complex learning, cognitive flexibility and attention (29). Knockout mice for Erbin (encoded by the mouse orthologue of *ERBB2IP*) show behavioral disturbances, including impairment in forebrain neural circuits and sensory gating and poor motor coordination (30). Synaptic transmission can also be modulated trans-synaptically, as in the case of the pre-synaptic adhesion molecule Neurexin-3 (encoded by the mouse orthologue of *NRXN3*), which can control the post-synaptic expression of the α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) glutamate receptors in the hippocampus (31).

All of these proteins are essential during the formation and maturation of synapses [Erbin (32), PSD-93/chapsyn-110 (33), Neurexin-3 (34)]. This evidence accrues on a robust body of evidence implicating synaptic function in ID and ASD pathophysiology (35). Further, some of the candidate genes we identified in the CNVs regulate brain development beyond synaptogenesis. For example, Astrotactin 2 (encoded by the mouse orthologue of *ASTN2*) is a membrane protein that mediates the formation of neuronal-glial adhesions during the migration of cerebellar granule neurons (36). Erbin is required for re-myelination of regenerated axons after injury (37). Interestingly, *ASTN2* and *SNX3* are also included in the endosome GO category, enriched in our analysis, supporting a known role for endocytosis mechanisms in neuron development and function (21).

Some of these genes encode for mRNAs that interact with the chromatin remodeler *CHD8* (*ERBB2IP*, *SNX3*) or with the translational regulator fragile X mental retardation protein (FMRP) (*DLG2*, *NRXN3*), suggesting that the synaptic expression and/or localization of these key synaptic components might undergo coordinated regulation.

Gene dosage alterations represent the most likely mechanism for the variants we identified, with most cases explained by haploinsufficiency of the candidate genes. It is also worth noticing that the

rearrangements in *DLG2* and *NRXN3* are transmitted by unaffected parents, indicating a variable expressivity and/or incomplete penetrance. This is in line with reports of intragenic deletions in *NRXN3* in individuals with ASD transmitted by parents with subclinical autistic traits or no clinical manifestations (38).

We also described two cases implicating regions susceptible to epigenetic dysregulation rather than altered gene dosage. One of our patients without skeletal anomalies had a *de novo* deletion overlapping the mesomelia-synostoses syndrome critical region (22). Since three cases with similar deletions in DECIPHER do not show sign of skeletal involvement, we formulate a hypothesis alternative to incomplete penetrance. We hypothesize that the deletions described by Isidor and colleagues disrupt a TAD boundary in one of the two genes (*SULF1*) they have proposed as haploinsufficient (Fig 3). We suggest that the ablation of this insulator element removes the constraints of a bone enhancer (likely acting on *SULF1* itself), allowing its action on genes located distally to *SULF1*. Based on the deletions reported in mesomelia-synostoses syndrome, a candidate gene is *NCOA2*, which encodes a transcriptional coactivator for nuclear hormone receptors (steroid, thyroid, retinoid, and vitamin D receptors) also known as TIF2. HEY1-NCOA2 fusion proteins have been found in specimens of mesenchymal chondrosarcoma, an aggressive, uncommon histologic tumor arising in bone and soft tissues (39). A similar mechanism has been proposed for another skeletal disorder, the mesomelic dysplasia Savarirayan type (MIM#645274). Three unrelated cases with 2-Mb overlapping *de novo* 6p22.3 microdeletions spanning four genes with no previous association with skeletal anomalies were identified (40). The discovery of a fourth individual with a larger deletion but no skeletal phenotype, as well as the identification of two TAD boundaries in the interval, suggest that the deletion might cause the aberrant activation of potential limb enhancers in the region (40).

In a second case, we found the deletion of a TAD boundary in proximity of *CTNND2* (Fig. 4). We propose that the deletion may cause altered expression of the encoded protein, δ-catenin, a regulator of synaptogenesis and dendritic morphogenesis cooperating with Erbin (32, 41). In fact, *Ctnnd2*

knockout mice exhibit structural and functional abnormalities of the synapses, as well as defects in spatial learning and fear conditioning (42). CNVs involving *CTNND2* have been described in individuals with ASD (24) and the gene has been proposed as an ASD risk gene based on rare mutations found in families with two or more severely affected females (24). Interestingly, the three cases in this family are female patients with psychological disturbances ranging from anorexia nervosa to anxiety disorder. This observation raises the intriguing possibility that loss-of-function or altered dosage of *CTNND2* might underlie DD/ID and/or ASD, while altered expression might result in a different phenotype.

Based on segregation analysis of a patient (case 296491) with an intragenic deletion of the *MBD5* gene, we classified deletions of exons 3-4 of the 5'UTR of this gene as likely benign. The deletion of exon 3 was shared with the father and paternal uncle, both unaffected (Fig. S2A). Further, DGV contains several unaffected controls carrying deletions in exons 3-4 (Fig. S2B). *MBD5* is considered a known disease gene (MIM#611472) and we think that reconsidering published cases with deletions in the 5'UTR may help to in further refine a consensus phenotype and explain cases of incomplete penetrance (43).

We also found six complex phenotypes with CNVs spanning a known disease gene or genomic disorder locus. These CNVs explain the clinical features only partially (Table S8). We think that these cases might be explained by compounding effects of the genomic disorder identified here and an independent Mendelian disorder, yet to be identified. Multiple molecular diagnoses have already been described in large surveys of patients analysed by diagnostic whole-exome sequencing, where they represent approximately 5% of solved cases (44).

In conclusion, clinical and genetic analysis of our cohort allowed identifying complex cases, possibly rare disease genes involved in DD/ID, and novel position effects.

FIGURE LEGENDS

Figure 1. Rationale for CNV interpretation.

The diagram summarizes the workflow and the findings of this study: 1,015 consecutive ID/DD cases were studied by 60K array-CGH. CNVs included in DGV as polymorphisms were considered benign, and the remaining CNVs were initially divided into pathogenic and VOUS, based on MIM# and literature data. Pathogenic CNVs were assigned to this category because considered large (> 3 Mb) and containing gene(s) relevant for the pathology, spanning known genomic disorders regions or known dosage sensitive disease genes. VOUS were divided into three categories exploiting several bioinformatics tools among which: Database of Genomics Variants (http://dgv.tcag.ca/dgv/app/home), DECIPHER, OMIM: Online Mendelian Inheritance in Man (www.omim.org). SFARI, NDAR, NDD genes and loci, ExAC, GO.

Figure 2. Evolutionary constraints and functional annotation of genes in likely pathogenic CNVs.

The scatter plot shows the UCSC genes spanned by the CNVs distributed based on their ExAC pLI (y-axis) and missense z-score (x-axis). Dotted lines represent thresholds for significant constraints. Relevant genes discussed in the text are indicated. Genes were also annotated based on the following functional categories: ASD risk genes (blue), DD risk genes (light blue), GWAS-associated schizophrenia *loci* (green), genes encoding synaptic proteins (orange), genes encoding PSD proteins (purple), genes encoding FMRP targets (gold), and genes encoding *CHD8* targets (beige).

Figure 3. TAD and chromosome 8q12.3q13.3 deletion.

Pedigree of case 314103 (arrow). Filled symbols: affected subjects; empty symbols: unaffected subjects; del: deletion; dup: duplication. The candidate gene is indicated in bold; blue bar: duplication;

red bar: deletion. On the right, deleted region in case 314103 and Hi-C profile derived from human ES cells associated to this region. Hi-C interactions are shown in a heatmap with each dot reflecting two interaction pairs. The resulting interaction profile shows the formation of triangles (schematically enhanced in color) that represent individual TADs. Dashed lines indicate predicted TAD boundaries. Red, deletion in our subject; grey, cases with an overlapping deletion reported in DECIPHER and published literature (22). The genes spanned in the 8q12.3q13.3 region between 62.5 and 72.5 Mb (hg19) are shown below. *SULF1* and *SLCO5A1*, genes associated with Mesomelia synostoses syndrome, and *EYA1* causing Branchiootic syndrome, are in bold. Below, H3K4Me1 and H3K27Ac chromatin signatures (indicative of active/poised enhancers) in the genomic region between 69.5 and 69.85 Mb.

Figure 4. TAD and chromosome 5p15.2 deletion.

Pedigree of case 300081 (see previous legend). On the right, deleted region in case 300081 and Hi-C profile derived from human ES cells, as described above. Dashed lines indicate predicted TAD boundaries. The red bar indicates the deletion in our subject. Genes located in 5p15.2 genomic region between 11.5 and 13.5 Mb (hg19) are shown, with *CTNND2,* associated with psychiatric disturbances, in bold. H3K4Me1 and H3K27Ac chromatin signatures in the genomic region between 11.5 and 13.5 Mb at chromosome 5p are shown.

REFERENCES

1. Miller DT, Adam MP, Aradhya S et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 2010: 86: 749-764.
2. van Bokhoven H. Genetic and epigenetic networks in intellectual disabilities. Annual review of genetics 2011: 45: 81-104.
3. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature 2015: 519: 223-228.
4. South ST, Lee C, Lamb AN et al. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genetics in medicine : official journal of the American College of Medical Genetics 2013: 15: 901-909.
5. Kearney HM, South ST, Wolff DJ et al. American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. Genetics in medicine : official journal of the American College of Medical Genetics 2011: 13: 676-679.
6. Di Gregorio E, Savin E, Biamino E et al. Large cryptic genomic rearrangements with apparently normal karyotypes detected by array-CGH. Mol Cytogenet 2014: 7: 82.
7. Biamino E, Di Gregorio E, Belligni EF et al. A novel 3q29 deletion associated with autism, intellectual disability, psychiatric disorders, and obesity. Am J Med Genet B Neuropsychiatr Genet 2016: 171: 290-299.
8. Belligni EF, Di Gregorio E, Biamino E et al. 790 Kb microduplication in chromosome band 17p13.1 associated with intellectual disability, afebrile seizures, dysmorphic features, diabetes, and hypothyroidism. Eur J Med Genet 2012: 55: 222-224.
9. Sanders SJ, He X, Willsey AJ et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 2015: 87: 1215-1233.
10. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014: 511: 421-427.
11. Zaidi S, Choi M, Wakimoto H et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature 2013: 498: 220-223.
12. Vervoort VS, Viljoen D, Smart R et al. Sorting nexin 3 (SNX3) is disrupted in a patient with a translocation t(6;13) (q21;q12) and microcephaly, microphthalmia, ectrodactyly, prognathism (MMEP) phenotype. Journal of medical genetics 2002: 39: 893-899.
13. Tarsitano M, Ceglia C, Novelli A et al. Microduplications in 22q11.2 and 8q22.1 associated with mild mental retardation and generalized overgrowth. Gene 2014: 536: 213-216.
14. Shaikh TH, O'Connor RJ, Pierpont ME et al. Low copy repeats mediate distal chromosome 22q11.2 deletions: sequence analysis predicts breakpoint mechanisms. Genome Res 2007: 17: 482-491.
15. Sahoo T, Theisen A, Rosenfeld JA et al. Copy number variants of schizophrenia susceptibility loci are associated with a spectrum of speech and developmental delays and behavior problems. Genetics in medicine : official journal of the American College of Medical Genetics 2011: 13: 868-880.
16. Noor A, Lionel AC, Cohen-Woods S et al. Copy number variant study of bipolar disorder in Canadian and UK populations implicates synaptic genes. Am J Med Genet B Neuropsychiatr Genet 2014: 165B: 303-313.
17. Glessner JT, Wang K, Cai G et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature 2009: 459: 569-573.

18. Giorgio E, Ciolfi A, Biamino E et al. Whole exome sequencing is necessary to clarify ID/DD cases with de novo copy number variants of uncertain significance: Two proof-of-concept examples. Am J Med Genet A 2016: 170: 1772-1779.
19. Van Den Bossche MJ, Strazisar M, Cammaerts S et al. Identification of rare copy number variants in high burden schizophrenia families. Am J Med Genet B Neuropsychiatr Genet 2013: 162B: 273-282.
20. Valenti D, de Bari L, de Filippis B et al. Mitochondrial dysfunction as a central actor in intellectual disability-related diseases: An overview of Down syndrome, autism, Fragile X and Rett syndrome. Neurosci Biobehav R 2014: 46: 202-217.
21. Patak J, Zhang-James Y, Faraone SV. Endosomal system genetics and autism spectrum disorders: A literature review. Neurosci Biobehav R 2016: 65: 95-112.
22. Isidor B, Pichon O, Redon R et al. Mesomelia-synostoses syndrome results from deletion of SULF1 and SLCO5A1 genes at 8q13. Am J Hum Genet 2010: 87: 95-100.
23. Hofmeister W, Nilsson D, Topa A et al. CTNND2-a candidate gene for reading problems and mild intellectual disability. J Med Genet 2015: 52: 111-122.
24. Turner TN, Sharma K, Oh EC et al. Loss of delta-catenin function in severe autism. Nature 2015: 520: 51-56.
25. Darnell JC, Van Driesche SJ, Zhang C et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 2011: 146: 247-261.
26. Cotney J, Muhle RA, Sanders SJ et al. The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat Commun 2015: 6: 6404.
27. Collins MO, Husi H, Yu L et al. Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. J Neurochem 2006: 97 Suppl 1: 16-23.
28. Bayes A, van de Lagemaat LN, Collins MO et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci 2011: 14: 19-21.
29. Nithianantharajah J, Komiyama NH, McKechanie A et al. Synaptic scaffold evolution generated components of vertebrate cognitive complexity. Nat Neurosci 2013: 16: 16-24.
30. Tao Y, Chen YJ, Shen C et al. Erbin interacts with TARP gamma-2 for surface expression of AMPA receptors in cortical interneurons. Nat Neurosci 2013: 16: 290-299.
31. Aoto J, Foldy C, Ilcus SM et al. Distinct circuit-dependent functions of presynaptic neurexin-3 at GABAergic and glutamatergic synapses. Nat Neurosci 2015: 18: 997-1007.
32. Arikkath J, Israely I, Tao Y et al. Erbin controls dendritic morphogenesis by regulating localization of delta-catenin. J Neurosci 2008: 28: 7047-7056.
33. Elias GM, Funke L, Stein V et al. Synapse-specific and developmentally regulated targeting of AMPA receptors by a family of MAGUK scaffolding proteins. Neuron 2006: 52: 307-320.
34. Dean C, Scholl FG, Choih J et al. Neurexin mediates the assembly of presynaptic terminals. Nat Neurosci 2003: 6: 708-716.
35. De Rubeis S, He X, Goldberg AP et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014: 515: 209-215.
36. Wilson PM, Fryer RH, Fang Y et al. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. J Neurosci 2010: 30: 8529-8540.
37. Liang C, Tao Y, Shen C et al. Erbin is required for myelination in regenerated axons after injury. J Neurosci 2012: 32: 15169-15180.
38. Vaags AK, Lionel AC, Sato D et al. Rare deletions at the neurexin 3 locus in autism spectrum disorder. Am J Hum Genet 2012: 90: 133-141.
39. Bishop MW, Somerville JM, Bahrami A et al. Mesenchymal Chondrosarcoma in Children and Young Adults: A Single Institution Retrospective Review. Sarcoma 2015: 2015: 608279.
40. Flottmann R, Wagner J, Kobus K et al. Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. J Med Genet 2015: 52: 476-483.

41. Arikkath J, Peng IF, Ng YG et al. Delta-catenin regulates spine and synapse morphogenesis and function in hippocampal neurons during development. J Neurosci 2009: 29: 5435-5442.
42. Matter C, Pribadi M, Liu X et al. Delta-catenin is required for the maintenance of neural structure and function in mature cortex in vivo. Neuron 2009: 64: 320-327.
43. Hodge JC, Mitchell E, Pillalamarri V et al. Disruption of MBD5 contributes to a spectrum of psychopathology and neurodevelopmental abnormalities. Mol Psychiatry 2014: 19: 368-379.
44. Posey JE, Harel T, Liu P et al. Resolution of Disease Phenotypes Resulting from Multilocus Genomic Variation. The New England journal of medicine 2017: 376: 21-31.

Table S1. CNVs >3Mb.

DECIPHER code	a-CGH (GRCh37/hg19)	Gender	Minimal region (Mb)	Inheritance	Protein coding genes	Syndrome	OMIM/ref.
301205	arr 1p13.3p22.1(93,688,983-109,445,775)x1~2	M	15.8	dn	56		
262739	arr 1q23.3q24.3(162,053,488-172,011,180)x1	M	10.1	dn	68		
283330*	arr 2p12p11.2(82,510,808-84,804,525)x1,14q11.2q12(20,472,548-31,139,579)x3,15q11.1q14(20,102,541-35,758,169)x1	M	2.2:10.7: 15.6	mat translocation	2:60:60		
296348	arr 2q37.1(231, 683,468-235,003,677)x3	M	3.3	dn	41		
296349	arr 2q37.2q37.3(235,875,302-243,041,364)x1	F	7.2	dn	62	Chr.2q37 deletion syndrome	600430
296352	arr 3q28q29(188,822,766 – 197,840,339)x3,9p24.3p23(271,257 – 13,003,711)x1	M	9:12.7	na	60:39	Chr.3q29 microduplication syndrome/Chr.9p deletion syndrome	611936/158170
295816	arr 4q31.22q31.3(147,830,359-151,242,375)x1	F	3.4	dn	8		
283327*	arr 4q34.1q35.2(172,930,618-190,896,674)x3, 9p24 3p23(271,257-12,907,826)x1,14q21 1(43.881,311-44,623,069)x3	M	18:12.6: 0.74	mat translocation	29:53:0		
296369	arr 4q34.2q35.1(177,498,192-185,093,872)x3.4q35.2(187,755,458-190,896,674)x3	F	7.6: 3.1	na	14: 4	Coats like	Reynolds et al., 2010
300747	arr 4q35.2(189,247,673-189,457,997)x3,Xp22.3q28(61,091-155,190,083)x3	F	0.210:155	mat: dn	0:812	47,XXX	
301208	arr 5q14.3q15 (84,194,538-97,051,910)x1	M	12.8	na	33		
296358	arr 6q13q14.1(73,228,354-82,363,752)x1	F	9.1	na	30		
296371	arr 6q22.31q22.33(121,604,976-129,216,265)x1	M	7.6	dn	27		
296373	arr 8p12q13.1(34,534,744-66,913,936)x3~2	M	32.4	na	110		
314103	arr 8q12.3q13.3(64,921,185-73,019,864x1)	F	8.1	dn	34		

300520	arr 9p24.3p13.1(271,257-39,156,954)x3	M	39	dn; mosaic iso9p	197		
306703	arr 10p15 3p14 (148,206-12,211,671) x1,12q24.31q24.33(12,572,578-133,767,986)x3	M	12:12	na	46:96		
283328*	arr 10p15.3p13(148,206-12,211,671)x1,12q24.31q24.33(121,572,578-133,767,986)x3	M	12:12.2	pat translocation	73:108		
300574	arr 10q22.3q23.2(81,581,543-88,816,677)x3	F	7.2	pat	28		
283660	arr 10q26.13q26.3(124,500,982-135,404,471)x1,12q24.31q24.33(125,178,836-133,819,092)x3	M	11:8.6	mat translocation	67:41		
296424	arr 1q23.3(160,932,162-160,978,746)x1,10q26.11q26.3(120,917,169-135,404,523)x1,22q13.33 (50,706,851-51,178,264)x3	M	0.05: 14.5: 0.5	mat translocation	1:88:20		
283326*	arr 11q24.3q25(128,728,456-134,868,407)x1,20q13.3(58,442,781-62,893,189)x3	F	6:4.4	dn	29:90	Jacobsen syndrome	147791
296404	arr 16q22.1q24.1(71,418,695-84,865,760)x3	M	13.4	na	85		
300038	arr 18q22.1q22.2(64,494,061-67,742,701)x3	M	3.2	mat	6		
301206	arr 20q13.13q13.33(49,620,164-58,379,837)x3	M	8.8	dn	47		
300572	arr 22q12.1q12.2(28,247,776-31,223,581)x1	M	3	dn	44	NF2	10100
296435	arr Xp22.33p11.1(2,700,316-57,932,274) x1, Xq11.1q28(58,051,706-154,841,455)x3	F	55.3:96.8	na	312:438		
301281	arr Xp11.22q28(61,091-155,232,907)x2,Yp11.32q12(11,091-59,335,913)x2	M	155.2:59.3	na	813:63	48,XXYY syndrome	

Notes. dn: *de novo*; mat: maternal; pat: paternal; na: not available. * Reported in (Di Gregorio et al., 2014)

Table S1. CNVs >3Mb.

DECIPHER code	a-CGH (GRCh37/hg19)	Gender	Minimal region (Mb)	Inheritance	Protein coding genes	Syndrome	OMIM/ref.
301205	arr 1p13.3p22.1(93,688,983-109,445,775)x1~2	M	15.8	dn	56		
262739	arr 1q23.3q24.3(162,053,488-172,011,180)x1	M	10.1	dn	68		
283330*	arr 2p12p11.2(82,510,808-84,804,525)x1,14q11.2q12(20,472,548-31,139,579)x3,15q11.1q14(20,102,541-35,758,169)x1	M	2.2:10.7:15.6	mat translocation	2:60:60		
296348	arr 2q37.1(231, 683,468-235,003,677)x3	M	3.3	dn	41		
296349	arr 2q37.2q37.3(235,875,302-243,041,364)x1	F	7.2	dn	62	Chr.2q37 deletion syndrome	600430
296352	arr 3q28q29(188,822,766 – 197,840,339)x3,9p24.3p23(271,257 – 13,003,711)x1	M	9:12.7	na	60:39	Chr.3q29 microduplication syndrome/Chr.9p deletion syndrome	611936/158170
295816	arr 4q31.22q31.3(147,830,359-151,242,375)x1	F	3.4	dn	8		
283327*	arr 4q34.1q35.2(172,930,618-190,896,674)x3, 9p24 3p23(271,257-12,907,826)x1,14q21 1(43.881,311-44,623,069)x3	M	18:12.6:0.74	mat translocation	29:53:0		

296369	arr 4q34.2q35.1(177,498,192-185,093,872)x3.4q35.2(187,755,458-190,896,674)x3	F	7.6: 3.1	na	14: 4	Coats like	Reynolds et al., 2010
300747	arr 4q35.2(189,247,673-189,457,997)x3,Xp22.3q28(61,091-155,190,083)x3	F	0.210:155	mat: dn	0:812	47,XXX	
301208	arr 5q14.3q15 (84,194,538-97,051,910)x1	M	12.8	na	33		
296358	arr 6q13q14.1(73,228,354-82,363,752)x1	F	9.1	na	30		
296371	arr 6q22.31q22.33(121,604,976-129,216,265)x1	M	7.6	dn	27		
296373	arr 8p12q13.1(34,534,744-66,913,936)x3~2	M	32.4	na	110		
314103	arr 8q12.3q13.3(64,921,185-73,019,864x1)	F	8.1	dn	34		
300520	arr 9p24.3p13.1(271,257-39,156,954)x3	M	39	dn; mosaic iso9p	197		
306703	arr 10p15 3p14 (148,206-12,211,671) x1,12q24.31q24.33(12,572,578-133,767,986)x3	M	12:12	na	46:96		
283328*	arr 10p15.3p13(148,206-12,211,671)x1,12q24.31q24.33(121,572,578-133,767,986)x3	M	12:12.2	pat translocation	73:108		
300574	arr 10q22.3q23.2(81,581,543-88,816,677)x3	F	7.2	pat	28		
283660	arr 10q26.13q26.3(124,500,982-135,404,471)x1,12q24.31q24.33(125,178,836-133,819,092)x3	M	11:8.6	mat translocation	67:41		

296424	arr 1q23.3(160,932,162-160,978,746)x1,10q26.11q26.3(120,917,169-135,404,523)x1,22q13.33(50,706,851-51,178,264)x3	M	0.05: 14.5: 0.5	mat translocation	1:88:20		
283326*	arr 11q24.3q25(128,728,456-134,868,407)x1,20q13.3(58,442,781-62,893,189)x3	F	6:4.4	dn	29:90	Jacobsen syndrome	147791
296404	arr 16q22.1q24.1(71,418,695-84,865,760)x3	M	13.4	na	85		
300038	arr 18q22.1q22.2(64,494,061-67,742,701)x3	M	3.2	mat	6		
301206	arr 20q13.13q13.33(49,620,164-58,379,837)x3	M	8.8	dn	47		
300572	arr 22q12.1q12.2(28,247,776-31,223,581)x1	M	3	dn	44	NF2	10100
296435	arr Xp22.33p11.1(2,700,316-57,932,274) x1, Xq11.1q28(58,051,706-154,841,455)x3	F	55.3:96.8	na	312:438		
301281	arr Xp11.22q28(61,091-155,232,907)x2,Yp11.32q12(11,091-59,335,913)x2	M	155.2:59.3	na	813:63	48,XXYY syndrome	

Notes. dn: *de novo*; mat: maternal; pat: paternal; na: not available. * Reported in (Di Gregorio et al., 2014)

Table S1. CNVs >3Mb.

DECIPHER code	a-CGH (GRCh37/hg19)	Gender	Minimal region (Mb)	Inheritance	Protein coding genes	Syndrome	OMIM/ref.
301205	arr 1p13.3p22.1(93,688,983-109,445,775)x1~2	M	15.8	dn	56		
262739	arr 1q23.3q24.3(162,053,488-172,011,180)x1	M	10.1	dn	68		
283330*	arr 2p12p11.2(82,510,808-84,804,525)x1,14q11.2q12(20,472,548-31,139,579)x3,15q11.1q14(20,102,541-35,758,169)x1	M	2.2:10.7: 15.6	mat translocation	2:60:60		
296348	arr 2q37.1(231, 683,468-235,003,677)x3	M	3.3	dn	41		
296349	arr 2q37.2q37.3(235,875,302-243,041,364)x1	F	7.2	dn	62	Chr.2q37 deletion syndrome	600430
296352	arr 3q28q29(188,822,766 – 197,840,339)x3,9p24.3p23(271,257 – 13,003,711)x1	M	9:12.7	na	60:39	Chr.3q29 microduplication syndrome/Chr.9p deletion syndrome	611936/158170
295816	arr 4q31.22q31.3(147,830,359-151,242,375)x1	F	3.4	dn	8		
283327*	arr 4q34.1q35.2(172,930,618-190,896,674)x3, 9p24 3p23(271,257-12,907,826)x1,14q21 1(43.881,311-44,623,069)x3	M	18:12.6: 0.74	mat translocation	29:53:0		
296369	arr 4q34.2q35.1(177,498,192-185,093,872)x3.4q35.2(187,755,458-190,896,674)x3	F	7.6: 3.1	na	14: 4	Coats like	Reynolds et al., 2010
300747	arr 4q35.2(189,247,673-189,457,997)x3,Xp22.3q28(61,091-155,190,083)x3	F	0.210:155	mat: dn	0:812	47,XXX	
301208	arr 5q14.3q15 (84,194,538-97,051,910)x1	M	12.8	na	33		
296358	arr 6q13q14.1(73,228,354-82,363,752)x1	F	9.1	na	30		
296371	arr 6q22.31q22.33(121,604,976-129,216,265)x1	M	7.6	dn	27		
296373	arr 8p12q13.1(34,534,744-66,913,936)x3~2	M	32.4	na	110		
314103	arr 8q12.3q13.3(64,921,185-73,019,864x1)	F	8.1	dn	34		

300520	arr 9p24.3p13.1(271,257-39,156,954)x3	M	39	dn; mosaic iso9p	197		
306703	arr 10p15 3p14 (148,206-12,211,671) x1,12q24.31q24.33(12,572,578-133,767,986)x3	M	12:12	na	46:96		
283328*	arr 10p15.3p13(148,206-12,211,671)x1,12q24.31q24.33(121,572,578-133,767,986)x3	M	12:12.2	pat translocation	73:108		
300574	arr 10q22.3q23.2(81,581,543-88,816,677)x3	F	7.2	pat	28		
283660	arr 10q26.13q26.3(124,500,982-135,404,471)x1,12q24.31q24.33(125,178,836-133,819,092)x3	M	11:8.6	mat translocation	67:41		
296424	arr 1q23.3(160,932,162-160,978,746)x1,10q26.11q26.3(120,917,169-135,404,523)x1,22q13.33 (50,706,851-51,178,264)x3	M	0.05: 14.5: 0.5	mat translocation	1:88:20		
283326*	arr 11q24.3q25(128,728,456-134,868,407)x1,20q13.3(58,442,781-62,893,189)x3	F	6:4.4	dn	29:90	Jacobsen syndrome	147791
296404	arr 16q22.1q24.1(71,418,695-84,865,760)x3	M	13.4	na	85		
300038	arr 18q22.1q22.2(64,494,061-67,742,701)x3	M	3.2	mat	6		
301206	arr 20q13.13q13.33(49,620,164-58,379,837)x3	M	8.8	dn	47		
300572	arr 22q12.1q12.2(28,247,776-31,223,581)x1	M	3	dn	44	NF2	10100
296435	arr Xp22.33p11.1(2,700,316-57,932,274) x1, Xq11.1q28(58,051,706-154,841,455)x3	F	55.3:96.8	na	312:438		
301281	arr Xp11.22q28(61,091-155,232,907)x2,Yp11.32q12(11,091-59,335,913)x2	M	155.2:59.3	na	813:63	48,XXYY syndrome	

Notes. dn: *de novo*; mat: maternal; pat: paternal; na: not available. * Reported in (Di Gregorio et al., 2014)

Table S4. Likely pathogenic CNVs.

Chr	DECIPHER code	Gender	Copy Number Variation	Size	Start min	End min	Protein coding genes	Inheritance	Candidate gene	Evidence from literature	Phenotype
					DE NOVO						
3	296509	M	dup3p21.1p14.3	1.1 Mb	53262296	54358357	7	dn	*CACNA1D*, *CACNA2D*, *DCP1A*		Global developmental delay
4	300731	M	dup4q31.1q31.21	1.6 Mb	140096835	141711184	13	dn	*NAA15*, *MAML3*, *NDUFC1*, *TBC1D9*		Facial dysmorphisms, Global developmental delay
4	296516 ”	M	del4q32.1	928 kb	157343163	158271008	3	dn	*GRIA2*		Cognitive impairment
6	296462 °”	F	del6q21	2.4 Mb	106727531	109121544	15	dn	*FOXO3*, *SNX3*, *BEND3*, *SEC63*, *TPD52L1*		Facial dysmorphisms, Cognitive impairment
14	296462 °”	F	del14q32.2	2.4 Mb	97636749	100053022	5	dn	*CCDC85C*, *BCL11B*		Facial dysmorphisms, Cognitive impairment

8	296542	M	dup8p23.1	634 kb	11207613	11841901	8	dn	*FDFT1*	Intellectual disability
9	296547	F	dup9p13.2	237 kb	37423984	37660586	6	dn	*GRHPR*	Facial dysmorphisms, Cognitive impairment
16	296562	M	dup16q23.1	383 kb	74872495	75255243	6	dn	*ZNFR1*	Cognitive impairment
19	299791	M	del19q12	1.68 Mb	29702580	31382941	8	dn	*UQCRFS1, ZNF536, POP4, URI1*	Autism, Behavioral abnormality, Severe intellectual disability, Seizures
19	300110	F	del19q13.42q13.43	664 kb	55689844	56353892	37	dn	*BRSK1, EPN1, FIZ1, PPP6R1, SYT5, ZNF524, ZNF579, ZNF580, ZNF581 ZNF628*	Global developmental delay
X	299788	F	dupXq21.1	690 kb	79368628	80059046	2	dn	*BRWD3*	Cognitive impairment
					INHERITED					
1	299799	F	dup1p34.1	113 kb	46650642	46764022	5	mat*	*LRRC41*	

										Facial dysmorphisms, Cognitive impairment, Proportionate short stature
1	300368	M	dup1q32.1	35 kb	202093675	202128833	3	mat	*ARL8A*	Global developmental delay
2	300355	M	del2p22.3	283 kb	38262957	38546186	3	pat	*ATL2*, *CYP1B1*	Facial dysmorphisms, Arnold-Chiari type I malformation, Facial capillary hemangioma, Macrocephaly, Thin corpus callosum
3	300203	M	dup3p26.3	2.1 Mb	93949	2185275	3	pat	*CHL, CNTN4*	Abnormal facial shape, Arachnoid cyst
4	300097	M	dup4p16.1	2.1 Mb	8281059	10417448	30	mat	*WDR1*	Cognitive impairment
4	299797	F	dup4p16.1	355 kb	9766686	10122009	3	pat*	*DRD5*, *WDR1*	Intellectual disability
4	265635	F	del4q21.23q22.1	1.7 Mb	86364433	88084471	7	pat	*MAPK10*	Intellectual disability, Microcephaly
4	300078 ’	M	dup4q21.3	197 kb	87080535	87277469	1	mat	*MAPK10*	Autism spectrum disorder
5	285092	F	del5q12.1	541 kb	61305070	61846296	3	mat*	*KIF2A*	Abnormal heart morphology, Congenital microcephaly, Delayed speech and language development, Global developmental delay,

										Hypothyroidism, Pulmonary hypoplasia, Short stature
5	263833	M	del5q12.3	466 kb	65081656	65547987	3	mat*	*ERBB2IP*	Abnormal emotion/affect behavior, Downslanted palpebral fissures, EEG with focal epileptiform discharges, Moderate intellectual disability, Seizures
7	300613	F	dup7p15.3	1.36 Mb	27433518	28790162	4	pat	*JAZF1*	Abnormality of the heart, Intellectual disability
8	283337 ’	F	dup8q13.3q21.11	3.8 Mb	72340784	76149340	16	pat	*RPL7*	Cognitive impairment, Delayed speech and language development
9	299884	F	del9q33.1	222 kb	119090300	119312402	2	mat*	*ASTN2*	Delayed speech and language development
10	299917	M	del10q24.1	357 kb	97373868	97731399	4	mat	*ENTPD1*	Cognitive impairment
11	300371	F	dup11q12.3	290 kb	62470205	62760655	16	mat	*RAB33B, SLC3A2, GNG3, HNRNPUL2, TTC9C*	Behavioural/Psychiatric Abnormality, Global developmental delay
11	300042	F	del11q14.1	373 kb	84046614	84419502	1	mat	*DLG2*	Deeply set eye, Hearing abnormality, Mild intellectual disability, Microcephaly,

										Myopia, Short digit, Synophrys
11	314659	M	del11q14.1	139 kb	84721282	84860632	1	pat	*DLG2*	Autism spectrum disorder
14	300101	M	del14q24.3	206 kb	77069559	77276003	2	pat	*VASH1*	Cognitive impairment, Lissencephaly
14	300039	F	del14q24.3q31.1	106 kb	79328112	79433755	1	mat	*NRXN3*	Global developmental delay
15	279524	F	dup15q26.3	266 kb	102044450	102310654	3	pat	*TARSL2, TM2D3*	Intellectual disability, severe, Macular hypopigmented whorls, streaks, and patches, Microcornea, Pulmonic stenosis, Unilateral microphthalmos
16	300202	F	del16p12.2	140 kb	21599687	21739885	3	pat	*METTL9*	Cognitive impairment
16	300046	M	dup16q23.3	1.4 Mb	82197554	83599577	2	mat	*CDH13*	Cognitive impairment
20	296418	M	del20q12	130 kb	41384436	41514592	1	mat	*PTPRT*	Facial dysmorphisms, Autism, Severe intellectual disability
22	266246	M	del22q11.21	247 kb	22323105	22569881	1	mat	*TOP3B*	Cerebellar vermis hypoplasia, Hypertelorism, Intellectual disability, Micrognathia, Synophrys, Tall stature, Triangular face
22	300526	F	del22q11.21	247 kb	22323105	22569881	1	mat*	*TOP3B*	Facial dysmorphisms, Autism spectrum disorder, Cognitive impairment

X	300040	M	delXq21.32q21.33	544 kb	92926160	93470236	2	mat	*NAP1L3*	Bilateral cryptorchidism, Cognitive impairment, Congenital bilateral hip dislocation, Seizures
Y	300214	M	delYp11.32	225 kb	1414171	1639610	4	pat	*SLC25A6*	Cognitive impairment
INHERITANCE NOT AVAILABLE										
1	299931	M	del1q23.1	1.75 kb	157103175	157104921	1	na	*ETV3*	Global developmental delay
1	300137 ”	F	dup1p31.1	483 kb	74954860	75438212	5	na	*CRYZ*	Intellectual disability
2	300321	M	dup2p23.1	165 kb	30301202	30466094	2	na	*YPEL5*	Autism spectrum disorder
6	300771	F	dup6p21.33	19 kb	31785578	31804641	3	na	*HSPA1B*	Generalized hypotonia, Global developmental delay, Linear hyperpigmentation
6	300151 ”	M	dup6q16.1	168 kb	96842882	97010537	2	na	*UFL1*	Generalized hypotonia, Global developmental delay, Seizures, Strabismus
7	296537	M	dup7q36.2	419 kb	152346292	152765153	2	na	*ACTR3B*	Intellectual disability
8	300157	M	dup8p23.1	787 kb	10358352	11145066	7	na	*SOX7*	Abnormal facial shape, Global developmental delay

8	300672 ”	M	dup8q11.23	227 Kb	53467337	53694718	2	na	*PMM2*	Abnormal facial shape, Cognitive impairment
9	300765	M	dup9q31.3	2.86 Mb	111341780	114198542	19	na	*PALM2*, *AKAP2*, *FRRS1L*, *TMEM245*	Abnormal facial shape, Cognitive impairment, Epilepsy, Global developmental delay
17	300149 ’	F	del17p13.2	129 kb	5235415	5364469	5	na	*C1QBP*, *RABEP1*	Cognitive impairment
12	300164	M	del12q24.33	53 kb	132475221	132528279	1	na	*EP400*	Cognitive impairment
15	300628	M	dup15q13.2q13.3	1.6 Mb	30938215	32510863	7	na	*KLF13*, *MTMR10*	Defect in the atrial septum, Global developmental delay
16	300770	M	dup16p12.2	406 kb	21973762	22380197	8	na	*UQCRC2*	Autism spectrum disorder
X	299932	F	dupXq22.1	589 kb	101751021	102339819	8	na	*GPRASP2*	Cognitive impairment, Global developmental delay, Intellectual disability, Seizures

Notes. dn: *de novo*; mat: maternal; pat: paternal; na: parents not available; *: parent affected with a similar or milder phenotype; °: subject with two likely pathogenic CNVs, see table S5. “: subjects with additional uncertain significance CNVs, see table S6;’: subjects with additional likely benign CNVs, see table S7.

Table S5. Candidate genes within the identified likely pathogenic CNVs.

DECIPHER code	Gene	Inheritance	Copy Number Variation	FMRP	G2C	Human Cortex PSD	SCZ loci	ASD TADA genes	DDG2P	CHD8	Evidence from literature	GO enrichment	pLi	Z
263833	*ERBB2IP*	mat*	del5q12.3		+	+		+			21170055	GO:0061245 establishment or maintenance of bipolar cell polarity biological_process	**0.99**	-1.33
265635	*MAPK10*	pat	del4q21.23q22.1		+				+		23329067	-	0.83	2.97
266246	*TOP3B*	pat	del22q11.21								24315824	-	0.11	**3.18**
279524	*TARSL2*	pat	dup15q26.3			+					-	-	0	-0.23
279524	*TM2D3*	pat	dup15q26.3							+	-	-	0	-0.69
283337	*RPL7*	pat	dup8q13.3q21.11			+					-	GO:0003729 mRNA binding molecular_function	**0.99**	1.58
283337	*STAU2*	pat	dup8q13.3q21.11									-	**0.95**	0.60
285092	*KIF2A*	mat*	del5q12.1		+	+			+		27896282	-	**1**	**4.44**
296418	*PTPRT*	mat	del20q12	+	+	+							**1**	**3.35**

296462	*BCL11B*	dn	del14q32.2					+			27762073	-	**0.93**	**6.42**
296462	*BEND3*	dn	del6q21							+	26481236	-	0.5	2.85
296462	*CCDC85C*	dn	del14q32.2							+	-	-	**-**	**-**
296462	*FOXO3*	dn	del6q21	+						+	-	-	**-**	**-**
296462	*NR2E1*	dn	del6q21								27434030	-	**0.99**	**3.54**
296462	*SEC63*	dn	del6q21							+	-	-	0	1.57
296462	*SNX3*	dn	del6q21		+	+				+	20817026	GO:0005768 endosome cellular_component GO:0010314 phosphatidylinositol-5-phosphate binding molecular_function	0.47	2.41
296462	*TPD52L1*	pat	dup6q22.31		+						-	-	0.01	1.05
296509	*CACNA1D*	dn	dup3p21.1p14.3						+		25620733	GO:1902495 transmembrane transporter complex cellular_component	**1**	**5.57**
296509	*CACNA2D3*	dn	dup3p21.1p14.3		+						-	-	**1**	0.95
296509	*DCP1A*	dn	dup3p21.1p14.3							+	19091970	-	**0.94**	-0.11

296516	*GRIA2*	dn	del4q32.1		+	+				22669415	GO:1902495 transmembrane transporter complex cellular_component	**1**	**4.43**
296537	*ACTR3B*	na	dup7q36.2		**+**					-	-	**1**	**3.84**
296542	*FDFT1*	dn	dup8p23.1						+	-	GO:0016765 transferase activity, transferring alkyl or aryl (other than methyl) groups molecular_function GO:0016491 oxidoreductase activity molecular_function	**-**	**-**
296547	*GRHPR*	dn	dup9p13.2			+				-	GO:0016491 oxidoreductase activity molecular_function	0	1.2
296562	*ZNFR1*	dn	dup16q23.1							26572622	-	0.75	1.57
299788	*BRWD3*	dn	dupXq21.1					+		24462886	-	**1**	**4.18**
299791	*POP4*	dn	del19q12		+					-	-	0.14	-0.17

299791	*UQCRFS1*	dn	del19q12				+				-	GO:0044455 mitochondrial membrane part cellular_component GO:0016491 oxidoreductase activity molecular_function GO:0016491 oxidoreductase activity molecular_function GO:1902495 transmembrane transporter complex cellular_component	0.69	2.47
299791	*URI1*	dn	del19q12						+		-	-	0.21	-1.25
299791	*ZNF536*	dn	del19q12	+				+			19398580	-	**0.97**	**3.27**
299797	*DRD5*	pat*	dup4p16.1								23505263	GO:0010578 regulation of adenylate cyclase activity involved in G-protein coupled receptor signaling pathway biological_process GO:0010579 positive regulation of adenylate cyclase activity involved in G-protein coupled receptor signaling pathway biological_process	0	0.56
299797	*WDR1*	pat*	dup4p16.1			+	+				-	-	**0.99**	1.7

299799	*LRRC41*	mat*	dup1p34.1	+						+	22709582	-	**1**	**3.18**
299884	*ASTN2*	mat*	del9q33.1		+						24381304	GO:0005768 endosome cellular_component	**0.99**	1.76
299917	*ENTPD1*	mat	del10q24.1						+		-	-	0.52	0.03
299931	*ETV3*	na	del1q23.1							+	-	-	0.78	1.57
299932	*GPRASP2*	na	dupXq22.1		+						-	-	0.5	-0.16
300039	*NRXN3*	mat	del14q24.3q31.1	+	+	+			+		26235839	-	**1**	**3.96**
300040	*NAP1L3*	mat	delXq21.32q21.33								-	-	0.41	0.33
300042	*DLG2*	mat	del11q14.1	+	+	+					27271353	-	0.67	1.64
300046	*CDH13*	mat	dup16q23.3		+	+					26460479	-	0.22	-1.03
300078	*MAPK10*	mat	dup4q21.3		+				+		23329067	-	0.83	2.97
300097	*WDR1*	mat	dup4p16.1		+	+					-	-	**0.99**	1.7
300101	*VASH1*	pat	del14q24.3							+	-	-	0.73	1.93
300110	*BRSK1*	dn	del19q13.42q13.43	+	+	+					24395778	-	**0.99**	**5.09**
300110	*EPN1*	dn	del19q13.42q13.43	+							-	-	0.43	1.49

300110	*FIZ1*	dn	del19q13.42q13.43					+	-	-	0.45	**5.05**
300110	*PPP6R1*	dn	del19q13.42q13.43					+	-	-	**1**	-0.35
300110	*SYT5*	dn	del19q13.42q13.43	+	+				-	GO:0005768 endosome cellular_component	0	1.32
300110	*ZNF524*	dn	del19q13.42q13.43					+	-	-	0.01	**1.14**
300110	*ZNF579*	dn	del19q13.42q13.43					+	-	-	-	-
300110	*ZNF580*	dn	del19q13.42q13.43					+	-	-	0.51	2.24
300110	*ZNF581*	dn	del19q13.42q13.43					+	-	-	0.06	1.9
300110	*ZNF628*	dn	del19q13.42q13.43	+					25329708	-	**0.99**	**5.05**
300137	*CRYZ*	na	dup1p31.1		+					-	0.00	-0.40
300149	*C1QBP*	na	del17p13.2	+	+				-	GO:0003729 mRNA binding molecular_function	0.07	0.52
300149	*RABEP1*	na	del17p13.2	+					-	-	**0.98**	1.27
300151	*UFL1*	na	dup6q16.1		+			+	-	-	0.00	-0.07
300157	*SOX7*	na	dup8p23.1	+					-	-	0.13	-0.08

300164	*EP400*	na	del12q24.33	+					+	25612291	-	**1**	**4.44**
300202	*METTL9*	pat	del16p12.2						+	-	-	0.38	2.5
300203	*CHL1*	pat	dup3p26.3			+				27346367	-	0	-3.86
300203	*CNTN4*	pat	dup3p26.3				+			27668389	-	**1**	-0.64
300214	*SLC25A6*	pat	delYp11.32			+				-	-	0.07	1.88
300321	*YPEL5*	na	dup2p23.1						+	-	-	0.8	2.7
300355	*ATL2*	pat	del2p22.3			+			+		-	**0.98**	0.23
300355	*CYP1B1*	pat	del2p22.3		+					-	GO:0097193 intrinsic apoptotic signaling pathway biological_process GO:0097267 omega-hydroxylase P450 pathway biological_process	0	0.38
300368	*ARL8A*	mat	dup1q32.1						+	-	-	0.63	**3.19**
300371	*GNG3*	mat	dup11q12.3			+				18956348	-	0.53	1
300371	*HNRNPUL2*	mat	dup11q12.3						+	-	-	**1**	2.38
300371	*RAB33B*	mat	dup11q12.3						+	-	-	0.9	1.5

300371	*SLC3A2*	mat	dup11q12.3				+					-	-	**0.92**	2.11
300371	*TTC9C*	mat	dup11q12.3								+	-	-	0	-0.39
300526	*TOP3B*	mat*	del22q11.21									24315824	-	0.11	**3.18**
300613	*JAZF1*	pat	dup7p15.3									-	-	**0.96**	3.06
300628	*KLF13*	na	dup15q13.2q13.3								+	27459725	-	0.63	**3.73**
300628	*MTMR10*	na	dup15q13.2q13.3								+	25370694	-	0.04	0.29
300672	*PMM2*	na	dup16p13.2								+	25596524	-	0	-2.14
300731	*MAML3*	dn	dup4q31.1q31.21								+	-	-	0.33	0.77
300731	*NAA15*	dn	dup4q31.1q31.21							+	+	25533962	-	**1**	**3.12**
300731	*NDUFC1*	dn	dup4q31.1q31.21								+	-	GO:0044455 mitochondrial membrane part cellular_component GO:0016491 oxidoreductase activity molecular_function	0.46	0.18
300731	*TBC1D*	dn	dup4q31.1q31.21	+							+	-	-	0.88	2.91
300765	*AKAP2*	na	dup9q31.3			+						-	-	0.07	-1.03

300765	*FRRS1L*	na	dup9q31.3				+		-	-	0.01	-0.46
300765	*PALM2*	na	dup9q31.3			+			-	-	0	0.35
300765	*TMEM245*	na	dup9q31.3			+		+	-	-	0	1.49
300770	*UQCRC2*	na	dup16p12.2		+	+		+	-	-	0.01	0.66
300771	*HSPA1B*	na	dup6p21.33		+	+			17568569	-	-	-
314659	*DLG2*	pat	del11q14.1	+	+	+			27271353	-	0.67	1.64

Notes. dn: *de novo*; mat: maternal; pat: paternal; na: parents not available; *: parent affected with a similar or milder phenotype

Table S6. Uncertain significance VOUS.

Chr	DECIPHER code	Gender	CNV	Size	Start min	End min	Protein coding genes	Inheritance
1	296487	M	dup1p32.3	176 kb	55075653	55252418	6	dn
1	300082	M	dup1q21.1	115 kb	145632334	145747269	3	na
1	300140 °	F	del1q44	229 kb	244527815	244756752	3	na
2	264646	F	del2q23	37 kb	111399243	111435995	1	mat
2	279468 °	F	del2p22.3	391 kb	33505098	33896081	3	pat
2	299975	F	del2p23.1	17 kb	30814684	30832151	1	mat
2	300074	F	dup2q11.2	42 kb	99743468	99786080	4	na
2	300108	M	del2q31.1	58 kb	175258350	175316315	3	pat
2	300160	M	del2p16.3	205 kb	49051215	49256330	1	na
2	300354 ¯	F	del2p13.3	55 kb	68846359	68901407	2	pat
2	301194	F	del2q13	203 kb	110841774	111044815	2	na
2	300105	M	dup2q13	203 kb	110841715	111044815	2	mat
2	300043	M	dup2q21.2q21.3	1.3 Mb	134095757	135406655	3	pat
3	299972	M	del3p26.3	487 kb	116412865	116899830	1	mat
3	299981	F	dup3q27.3	196 kb	186598363	186794813	1	mat
3	300100	M	del3p22.2	154 kb	37366472	37520181	2	mat
3	300159	F	del3p12.3	93 kb	76631976	76724726	1	pat
3	300213	M	del3p12.3	82 kb	76549883	76632035	1	na
3	300216	M	del3q26.32	78 kb	178742237	178820723	1	na
3	300372	M	del3p14.2	1.7 kb	59956919	59958621	1	pat
3	300768 °	M	del3p14.1	649 kb	68905749	69554705	7	na

4	300139	M	dup4p14	259 kb	38858095	39117341	5	na
4	300147	F	dup4q12	198 kb	57082664	57280594	3	na
4	300374	M	dup4q12	398 kb	54766016	55164299	4	pat
4	301199	M	dup4p15.33	12 kb	17488306	17500286	1	na
4	296516 “	M	del4q34.1	219 kb	174941276	175160361	1	dn
5	286483	F	dup5p15.2	1.8 Mb	11843306	13266696	1	dn
5	296528	F	dup5q35.3	197 kb	179878423	180075503	3	dn
5	300102	M	dup5q21.1	174 kb	59739027	59913155	2	na
5	300119	M	dup5q27.1q27.2	190 kb	178727791	178917594	1	na
5	300166	F	dup5q14.1	279 kb	78280046	78559013	5	pat
5	300209	M	dup5q31.1q31.2	878 kb	135790418	136668688	1	na
5	300773	M	dup5q32	568 kb	146236055	146804546	3	na
5	301268	F	dup5q14.3	1.3 Mb	83561124	84869011	1	na
5	300081 °	F	del5p15.2	284 kb	12464932	12748960	0	mat
6	301218 °	M	del6p22.1	74 kb	26428923	26503116	3	na
6	272293	M	dup6q15	196 kb	88170758	88366616	4	mat
6	296462 ”	F	dup6q22.31	575 kb	125083046	125657899	4	pat
6	282113	M	dup6q26	1 Mb	161951066	162976601	1	mat
6	300208	M	dup6q11.1q12	818 kb	63197936	64016407	1	na
6	300772	M	del6q26	509 kb	162541352	163050403	1	mat
7	301269	F	dup7p22.3	375 kb	995544	1370682	7	na
8	283767	M	del8q21.3	100 kb	91611770	91712507	1	pat
8	300106	M	dup8q22.3	279 kb	104220546	104499653	5	pat
8	300151 ”	M	dup8q11.21	96 kb	48842451	48938844	3	mat
9	300137 ”	F	del9p24.1	275 kb	6055497	6330726	2	na

9	299887	M	del9q22.32	106 kb	96827056	96933017	1	pat
9	299980	M	del9p24.1	126 kb	3098538	3224927	1	na
9	300115	M	del9p22.2	529 kb	16973983	17502944	1	na
9	300134	F	dup9q32	230 kb	115379506	115609412	3	mat
9	300138	M	dup9q21.32q21.33	238 kb	86801352	87039515	1	pat
10	296553	F	del10p11.22	134 kb	32095083	32229198	1	dn
10	296557	F	dup10q23.1	99 kb	82172109	82270856	2	na
10	300079	M	dup10q25.3	309 kb	117723925	118032756	1	mat
10	300081 °	F	dup10q11.2q11.1	643 kb	47011584	47655146	2	na
10	300098	M	del10q23.33	159 kb	96453830	96612764	2	mat
10	300104	M	dup10p11.21	266 kb	35061767	35327968	2	mat
11	279468 °	F	dup11p14.2p14.1	391 kb	26834356	27225374	2	pat
11	300109	M	del11q14.1	162 kb	84419443	84581292	1	mat
11	300111	M	del11q14.1	172 kb	84367238	84539665	1	mat
12	282690	F	del12p12.2p12.1	345 kb	21032465	21377322	3	pat
12	299914	M	dup12q13.12	68 kb	49742919	49811367	2	na
12	300072	M	del12p12.2p12.1	345 kb	21032465	21377322	3	na
12	300164	M	del12q24.33	53 kb	132475221	132528279	1	na
12	300204	F	del12q13.13	31 kb	53806460	53837713	4	na
12	300356	M	dup12p13.2	629 kb	10242688	10872197	14	na
12	300370	M	dup12q13.2	38 kb	57336270	57374466	1	mat
12	300769	F	dup12q24.33	135 kb	132475221	132610696	1	na
13	300150 °	M	dup13q14.11	187 kb	43544758	43731917	2	na
15	299892	F	del15q22.31	13 kb	66828787	66841841	2	na
15	299913	F	del15q23	85 kb	69192894	69277766	2	mat

15	300162	M	dup15q26.3	373 kb	99244988	99618265	2	mat
15	301274	M	dup15q21.3	1.1 Mb	54757926	55837577	8	na
15	299973	M	dup15q13.3	489 kb	32021733	32510863	2	na
15	300534	M	dup15q13.3	489 kb	32021733	32510863	2	pat
15	300212	F	dup15q13.3	483 kb	32027733	32510863	2	mat
15	300527	M	dup15q13.3	417 kb	32021733	32438943	2	pat
15	301271	M	dup15q13.3	489 kb	32021733	32510863	2	pat
16	266337	F	del16q12.2	13 kb	56672771	56685867	2	mat
16	300099	M	del16q23.1	138 kb	78586260	78724670	1	pat
16	300117	M	dup16p13.12p13.11	234 kb	14545934	14780194	3	mat
16	300135 ¯	F	del16p13.3	127 kb	6881091	7008383	1	na
16	300140 °	F	del16q23.3	253 kb	82197554	82450270	1	na
16	300150 °	M	del16p13.11	23 kb	15131723	15154746	3	na
16	300672 ”	M	dup16p13.2	137 kb	8878007	9015169	5	na
17	300073	F	dup17p13.2	171 kb	5388734	5560010	3	pat
17	300207	M	del17p11.2	68 kb	21433385	21501929	1	na
19	299786	M	del19p12	418 kb	23624728	24043055	2	na
19	299888	M	dup19p13.2	125 kb	9678768	9803405	3	mat
19	299930	M	dup19q13.12	129 kb	37553808	37683194	3	dn
19	299969	M	del19p12	56 kb	23855084	23911266	1	na
19	300041 ¯	M	dup19p13.2	148 kb	13327639	13476228	1	na
19	300768 °	M	del19p12	418 kb	23624728	24043055	2	na
19	301218 °	M	del19p12	137 kb	23855084	23991913	2	na
20	299880 ¯	M	del20p11.21	124 kb	23421007	23545476	4	pat
21	300373	F	dup21q11.2	353 kb	15372452	15725877	2	na

21	301183	F	dup21q21.1	217 kb	18819200	19036035	2	na
22	299790	M	dup22q13.2	302 kb	40139760	40441961	4	dn
X	299920	M	dupXq21.1	507 kb	78367005	78873820	2	mat
X	299921	M	dupXp22.2	908 kb	13640036	14548099	8	mat
X	299789	M	dupXp22.31	1.56 Mb	6552712	8115153	4	mat
X	300165 ¯	F	dupXp22.31	1.56 Mb	6552712	8115153	4	dn
X	300211	F	delXp22.31	1.56 Mb	6552712	8115153	4	mat
X	300318	M	dupXp22.31	328 kb	6705268	7033316	1	na
X	300320	M	dupXp11.2q11.1	733 kb	57318899	58051765	3	na
X	300766	M	dupXq12	290 kb	65815490	66105563	1	na
X	301185	M	dupXp22.3	520 kb	998108	1518259	4	mat
X	301187	M	dupXp22.33	98 kb	1217017	11314792	2	mat
Y	300158	M	dupYp11.2	341 kb	6716801	7057535	2	na

Notes. dn: *de novo*; mat: maternal origin; pat: paternal origin; na: inheritance not available; °: subjects with two uncertain significance CNVs; “: subjects with additional likely pathogenic CNVs, see table S4; ¯: subjects with additional likely benign CNVs.

Table S7. Likely benign VOUS.

Chr	DECIPHER code	Gender	CNV	Size	Start min	End min	Protein coding genes	Inheritance
1	296482	F	del1p31.1	82 kb	73718381	73800673	0	na
1	299918	M	dup1q44	450 kb	246176039	246626354	1	pat
2	300071	F	del2p16.3	134 kb	52434634	52569208	0	pat
2	300107	M	del2p16.1	810 kb	56883079	57692707	0	na
2	300118	F	dup2p25.3	443 kb	708239	1151213	1	pat
2	300133	M	del2p11.2p11.1	792 kb	89441848	90234023	0	mat
2	300136	M	del2p22.3	94 kb	35414006	35508026	0	na
2	300369	M	dup2q14.2	186 kb	121600293	121786343	1	mat
2	300764	M	del2p22.3	144 kb	35507967	35652568	0	na
2	296491	M	del2q23.1	122 kb	148879622	149002494	1	pat
3	299885	F	dup3q26.2	719 kb	167837277	168556684	0	mat
3	300047 °	M	dup3q21.3	214 kb	126302115	126516322	3	mat
3	300156	M	dup3q23q24	632 kb	142840205	143472170	2	pat
3	300206	M	dup3p12.2	585 kb	80616137	81201574	0	na
4	299798	M	del4q13.3	112 kb	71162798	71275324	4	mat
4	299923 °	M	del4q13.3	112 kb	71162798	71275324	4	mat
5	300205	M	dup5q23.1	70 kb	115508329	115578519	1	mat
7	282534	M	dup7p21.2	700 kb	14849569	15548854	2	mat
7	300080	M	del7q21.3	12 kb	97483869	97495857	1	pat
7	300210	F	dup7p14.3	288 kb	29231988	29519980	2	pat
7	300078 ’	M	dup7q31.33	80 kb	123952457	124029497	0	na
8	300354 ¯	F	dup8q13.2	99 kb	68026085	68125318	2	mat
8	300375	M	del8q21.13	588 kb	83182853	83771019	0	mat
9	299801	F	dup9p23	521 kb	10060015	10581502	1	pat
10	300047 °	M	del10q25.1	98 kb	111304044	111401818	0	na
10	300215 °	M	del10q21.3	164 kb	65621077	65784648	0	mat
12	300149 ’	F	dup12p12.1	279 kb	24985765	25264854	3	na
12	300135 ¯	F	dup12q24.33	135 kb	132475221	132610696	1	pat
13	283337 ’	F	del13q21.33	692 kb	70370848	71062601	1	pat
15	300163	F	dup15q26.1q26.2	898 kb	93900718	94798487	1	pat
16	300165 ¯	F	dup16q24.1	180 kb	85942729	86122957	1	mat
17	266402	F	dup17p13.3	39 kb	183662	223241	1	pat
19	300215 °	M	dup19q13.12q13.13	164 kb	38229263	38393630	2	mat
21	300041 ¯	M	del21q21.1q21.2	241 kb	23685575	23927226	0	mat
22	299970	M	dup22q13.32	276 kb	49371757	49647323	0	na

X	299880 ¯	M	dupXp11.4p11.3	689 kb	41841241	42529887	0	mat
X	299886	M	dupXq26.3q27.1	184 kb	136957262	137140970	0	mat
X	299928	F	dupXq21.32	549 kb	92377119	92926219	1	pat
X	300103	F	dupXp22.2	663 kb	11258224	11921646	3	mat
X	300317	M	dupXq21.2q21.31	566 kb	86144038	86709812	0	mat
X	300767	M	dupXp11.2	177 kb	53727695	53904567	0	mat
X	301179	F	dupXq27.1	186 kb	138826286	139012778	1	mat
Y	299923 °	M	dupYq11.21	454 kb	14324324	14777373	0	pat

Notes. This table lists all the likely benign CNVs identified in our analysis (n: 43). Among them only 32 cases were counted in fig. 1. The remaining subjects were included in the other two VOUS groups, having more than one CNV.

dn: de novo; mat: maternal origin; pat: paternal origin; na: inheritance not available; °: subjects with two likely benign CNVs. ’: subjects with additional likely pathogenic CNVs, see table S4; ¯: subjects with additional uncertain significance CNVs.

Table S8. Six cases with complex phenotypes.

DECIPHER code	a-CGH (GRCh37/hg19)	Minimal region	Inheritance	Genes involved	Gene/Syndrome	OMIM/ ref.	Phenotype associated with the CNV	Extra-phenotype	Group
300534	arr 15q13.3(32,021,733-32,510,863)x3	489 kb	pat	2	*CHRNA7*	118511	Cognitive impairment, Neonatal hypotonia, Patent foramen ovale	Cerebellar vermis hypoplasia, Pontocerebellar hypoplasia	Uncertain VOUS, p
300533	arr 1p36.21(15,443,521-15,739,333)x3, 16p13.12p12.3(15,048,751-16,276,115)x1	30 kb: 1.2 Mb	mat*	6: 20	Chr. 16p13.11 deletion	Tropeano et al. (2013)	Cognitive impairment, Seizures	Generalized ichthyosis, Congenital ichthyosiform erythroderma (collodion baby)	Likely be Pathogen
300525	arr 15q11.2(22,784,523-23,179,948)x1, 17q21.2(39,406,803-39,645,565)x3	395 kb: 239 kb	mat	5: 22	Chr. 15q11.2 deletion	615656	Absence seizures, Autism spectrum disorder, Global developmental delay, Sagittal craniosynostosis	Postaxial feet polysyndactyly	Pathogen Likely be

300606	arr 9p24.3p23(611,628-8,842,561)x1	8.2 Mb	na	77	Chr. 9p deletion	158170	Global developmental delay, Single umbilical artery, Trigonocephaly, Ventricular septal defect	Abnormality of the nasolacrimal system (eye fistula)	Pathogen
300741	arr 22q11.21(20,754,422-21,440,514x1)	686 kb	na	30	Chr. 22q11.2 deletion	611867	Generalized hypotonia, Intellectual disability, Ligamentous laxity, Nystagmus	Hypopigmentation of the fundus, Leber's congenital amaurosis	Pathogen
296601	arr 18q12.2(32,883,148-35,329,908)x1	2.5 Mb	na	19	Chr. 18q12.2 deletion	Halgren et al. (2012)	Abnormal facial shape, Behavioural/Psychiatric Abnormality, Intellectual disability	Corpus callosum lipoma	Pathogen

Notes. mat: maternal origin; pat: paternal origin; na: inheritance not available; *: parent affected with a similar or milder phenotype.

Supporting information

METHODS

Cases

1,015 consecutive cases diagnosed with DD/ID in the Medical Genetics Unit at the "Città della Salute e della Scienza" University Hospital, Turin, Italy, over a 7-year period (2008-2014) were analysed with array-CGH. Cases were 1-17 years old and FRAXA was excluded by genetic test. Genomic DNA was extracted from peripheral blood (Qiagen, Hilden, Germany), and quality assessed using Nanodrop spectrophotometer (Thermo Scientific, Waltham, Massachusetts, USA). Array-CGH was performed using a 60K whole-genome oligonucleotide microarray (SurePrint G3 Human CGH Microarray 8x60K) following the manufacturer's protocol (Agilent Technologies, Santa Clara, California, USA). The overall median probe spacing is 41 kb (33 kb in RefSeq genes); the number of RefSeq genes covered is 15,553 (>1 probe, 83.2%) and 4,580 (≥3, 24.5%) (http://www.agilent.com/cs/library/brochures/5990-3368en_lo.pdf). Slides were scanned using a G2565BA device, and analyzed using Agilent CGH Analytics software ver. 4.0.81 (Agilent Technologies Inc.) with the statistical algorithm ADM-2 and a sensitivity threshold of 6.0. At least three consecutive aberrant probes identified significant copy number changes. Reference genome was the GRCh37/hg19 assembly. Pathogenic or potentially pathogenic CNVs identified by array-CGH were validated by quantitative real-time polymerase chain reaction (qPCR; primer sequences and conditions available upon request). We verified segregation whenever parental DNA was available. Written informed consent was obtained from all family members or their legal guardian, and IRB approved the study.

Pathogenicity analyses

For the interpretation of pathogenicity, we integrated several tools. We compared the variants identified in our cohort to those identified in the general population reported in the DGV. We also considered DECIPHER, the Online Mendelian Inheritance in Man (OMIM, http://www.omim.org/), the Human Gene Mutation Database (HGMD, http://www.biobase-international.com/product/hgmd), Simons Foundation Autism Research Initiative (SFARI, https://sfari.org), National Database for Autism Research (NDAR, https://ndar.nih.gov) and literature mining.

Additionally, we scanned CNVs for the presence of genes and *loci* recently associated to neurodevelopmental and co-morbid psychiatric conditions. We extracted the list of 65 ASD risk genes from (1), the list of 108 genome-wide significant *loci* in schizophrenia from the (2), and the list of 1,175 genes of brain-associated genes of "DD gene-disease pairs and attributes" from the Gene2Phenotype database (http://www.ebi.ac.uk/gene2phenotype/gene2phenotype-webcode/cgi-bin/handler.cgi). According to (3), VOUS were further classified as likely pathogenic, of unknown significance and likely benign. The VOUS not assigned to the "likely pathogenic category" were assigned to the "likely benign" group if they did not include any gene, or involved genes unlikely to be haploinsufficient (HI score >90%, (4)), or were partial duplications of a gene, inherited from an unaffected parent. The remaining VOUS were included to the "uncertain significance" group.

Subjects with more than one CNV were counted only once, considering them in the most "deleterious" group of pathogenicity analysis (gradient of decreasing severity from left to right in the last row of Fig. 1)

Overlap with functional categories relevant to neurodevelopmental disorders

We considered the probability of intolerance to loss-of-function (LoF) variants (pLI) and the deviation of the observed number of missense variants from the expected number (Mis z-score), as computed by the Exome Aggregation Consortium (ExAC) (http://exac.broadinstitute.org/) (5). Z-scores of 3.09 or higher indicate intolerance to missense variation. Genes with pLI scores of 0.9 or higher are extremely intolerant to heterozygous LoF variation, and thus haploinsufficient (5). We also considered functional categories relevant to neurodevelopmental risk: i) genes encoding mRNAs that interact with the RNA-binding protein FMRP, which is absent or mutated in Fragile X syndrome and a major hub for the regulation of ASD genes (6); ii) genes targeted by the chromatin remodeler *CHD8*, a prominent ASD gene (7) acting as a regulatory hub for other ASD genes (8); iii) human orthologues of mouse genes encoding for synaptic proteins, as reported in the Genes2Cognition (G2C) database (http://www.genes2cognition.org/) (9); iv) human genes encoding proteins of the postsynaptic density (PSD) isolated from postmortem human neocortex (10) (Fig. 2; Tables S4 and S5) . For each subject, we selected the most likely candidate gene based on: i) genetic evidence as an ASD, DD or schizophrenia genes, compatible with inheritance pattern (as defined above, recessive genes disregarded) or robust evidence from the literature (see Table S5); ii) pLI and/or Mis z-score constraints and at least a functional evidence, as described above; iii) prioritization of candidate genes within an interval based on i) and ii) (Fig. 2).

The 842 mRNAs targeted by FMRP were extracted as in (11). The target genes by *CHD8* in human neural stem cells, midfetal human brain and embryonic (E17.5) mouse cortex were extracted as in (8) (Supplementary Data 1). Human orthologues of mouse genes encoding synaptosomal proteins were extracted from Genes2Cognition database (http://www.genes2cognition.org, lists L00000009 to L00000016) and are based on (9). Human genes encoding PSD proteins were extracted from (10).

Constraint scores

Constraint metrics were obtained from the Exome Aggregation Consortium (ExAC, Cambridge, MA; URL: http://exac.broadinstitute.org) [May 2016]. We used the missense Z score, computed as deviation of observed number of missense variants from the expected number, as a measure of intolerance to missense variation (12). We used the probability of loss-of-function intolerance (pLI) as a measure of intolerance to loss-of-function mutations (12).

Gene Ontology (GO) analysis

Gene Ontology (GO) analysis was conducted on the CNVs with a size ≤ 5 Mb. Overlapping regions were considered only once. We obtained the list of genes included in 111 pathogenic, 52 likely pathogenic, 109 uncertain significance and 43 likely benign regions, according to the hg19 human genome release. We calculated statistical significance of GO enrichments using an empirical pValue. We compared the number of regions where at least one gene related to the GO found, to the distribution of regions with the same hit obtained from 1,000 randomizations of the regions set (Bioconductor RegioneR, http://bioconductor.org/packages/release/bioc/html/regioneR.html). We considered GO terms with ≤ 2000-associated genes and an empirical pValue ≤0.01. We further ranked significant GO keywords according to their corresponding number of related regions, and retained those ≥2.

Topological domains analysis

Topological domain data from genome-wide higher order chromatin interaction data in human embryonic stem cells (13) were downloaded from http://chromosome.sdsc.edu/mouse/hi-

c/download.html and mapped to hg19 coordinates using the UCSC LiftOver tool. Topological associated domain (TAD) boundaries are defined as regions with size up to 400,000 base pairs (400 kb) between topological domain regions.

DNase-seq reads from the NIH Roadmap Epigenomics Mapping Consortium were downloaded (14) and tissue-specific enhancers were predicted on the basis of differential DNase I hypersensitivity profiles from 10 human tissues according to Ibn-Salem et al. (15).

Figure S1. Likely pathogenic variants in *ERBB2IP* and *NAA15*.

A) Pedigree of case 263833 showing a 455 kb deletion at chromosome 5q12.3 shared by the proband and his affected mother. The deletion spans three genes: *NLN*, *ERBB2IP* and *SREK1*. *ERBB2IP* is suggested as the candidate within the interval. B) Family pedigree (case 300731) of the case with the 1.6 Mb *de novo* duplication of 4q31.1q31.21. Amongst duplicated genes, *NAA15* is suggested as the candidate gene for the patient's phenotype. In DECIPHER database one subject with ID and a duplication encompassing *NAA15* is reported. Family pedigrees: filled symbols - affected subjects; empty symbols - unaffected subjects; arrow indicates the proband; del: deletion; dup: duplication. The candidate gene is indicated in bold; blue bar: duplication; red bar: deletion.

Figure S2. Deletion in the 5'UTR of *MBD5*.

A) Family pedigree (case 296491) showing a partial *MBD5* deletion that is shared by the proband, his father and paternal uncle, both unaffected. Symbols as in the previous legend. B) The deletion spans exons 3 and 4, encoding part of the 5'UTR of *MBD5*. Three cases spanning exons

3-5 of the gene (also located in the 5'UTR) reported in control subjects in DGV are indicated in grey.

References

1. Sanders SJ, He X, Willsey AJ et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 2015: 87: 1215-1233.
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. Nature 2014: 511: 421-427.
3. Kearney HM, South ST, Wolff DJ et al. American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities. Genetics in medicine : official journal of the American College of Medical Genetics 2011: 13: 676-679.
4. Huang N, Lee I, Marcotte EM et al. Characterising and predicting haploinsufficiency in the human genome. PLoS Genet 2010: 6: e1001154.
5. Lek M, Karczewski KJ, Minikel EV et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016: 536: 285-291.
6. De Rubeis S, He X, Goldberg AP et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014: 515: 209-215.
7. Neale BM, Kou Y, Liu L et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012: 485: 242-245.
8. Cotney J, Muhle RA, Sanders SJ et al. The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. Nat Commun 2015: 6: 6404.
9. Collins MO, Husi H, Yu L et al. Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. J Neurochem 2006: 97 Suppl 1: 16-23.
10. Bayes A, van de Lagemaat LN, Collins MO et al. Characterization of the proteome, diseases and evolution of the human postsynaptic density. Nat Neurosci 2011: 14: 19-21.
11. Darnell JC, Van Driesche SJ, Zhang C et al. FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. Cell 2011: 146: 247-261.
12. Lek M, Karczewski K, Minikel E et al. Analysis of protein-coding genetic variation in 60,706 humans. bioRxiv 2016.
13. Dixon JR, Selvaraj S, Yue F et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012: 485: 376-380.
14. Bernstein BE, Stamatoyannopoulos JA, Costello JF et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol 2010: 28: 1045-1048.
15. Ibn-Salem J, Kohler S, Love MI et al. Deletions of chromosomal regulatory boundaries are associated with congenital disease. Genome biology 2014: 15: 423.

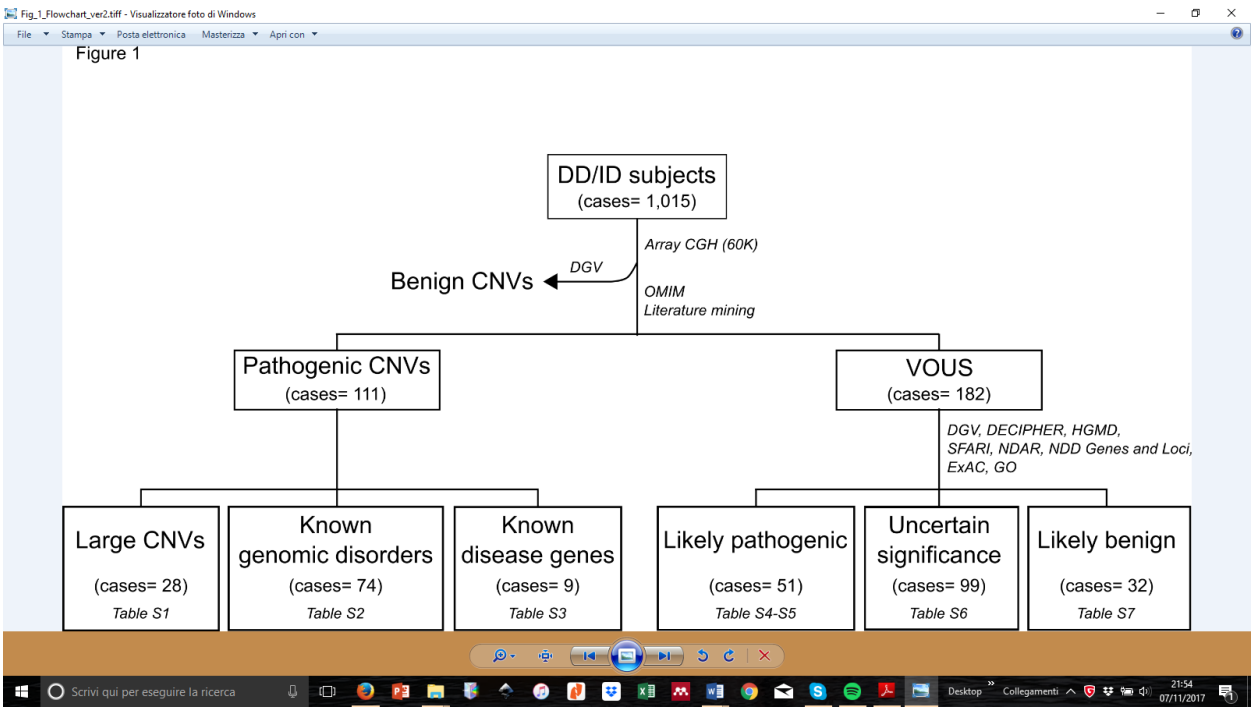


Fig_1_Flowchart_ver2.tiff - Visualizzatore foto di Windows
File
Stampa
Posta elettronica
Masterizza
Apri con
Figure 1
DD/ID subjects
(cases= 1,015)
Array CGH (60K)
DGV
Benign CNVs
OMIM
Literature mining
Pathogenic CNVs
(cases= 111)
VOUS
(cases= 182)
DGV, DECIPHER, HGMD,
SFARI, NDAR, NDD Genes and Loci,
ExAC, GO
Large CNVs
(cases= 28)
Table S1
Known
genomic disorders
(cases= 74)
Table S2
Known
disease genes
(cases= 9)
Table S3
Likely pathogenic
(cases= 51)
Table S4-S5
Uncertain
significance
(cases= 99)
Table S6
Likely benign
(cases= 32)
Table S7
Scrivi qui per eseguire la ricerca
Desktop
Collegamenti
21:54
07/11/2017


Fig_2_CNVs_highlighted.tiff - Visualizzatore foto di Windows
File
Stampa
Posta elettronica
Masterizza
Apri con
Figure 2
ASD genes
DD genes
SCZ loci
SYN genes
PSD genes
FMRP target genes
CHD8 target genes
LRRC41
NRXN3
PTPRT
EP400
CACNA1D
NAA15
KIF2A
BRSK1
CNTN4
CACNA2D3
WDR1
BRWD3
ERBB2IP
ATL2
STAU2
ASTN2
ZNF536
ACTR3B
GRIA2
ZNF628
BCL11B
MAPK10
DLG2
SNX3
TOP3B
DRD5
pLI
Mis z score
0.9
0.6
0.3
0.0
−5.0
−2.5
0.0
2.5
5.0
7.5
Scrivi qui per eseguire la ricerca
Desktop
Collegamenti
21:55
07/11/2017

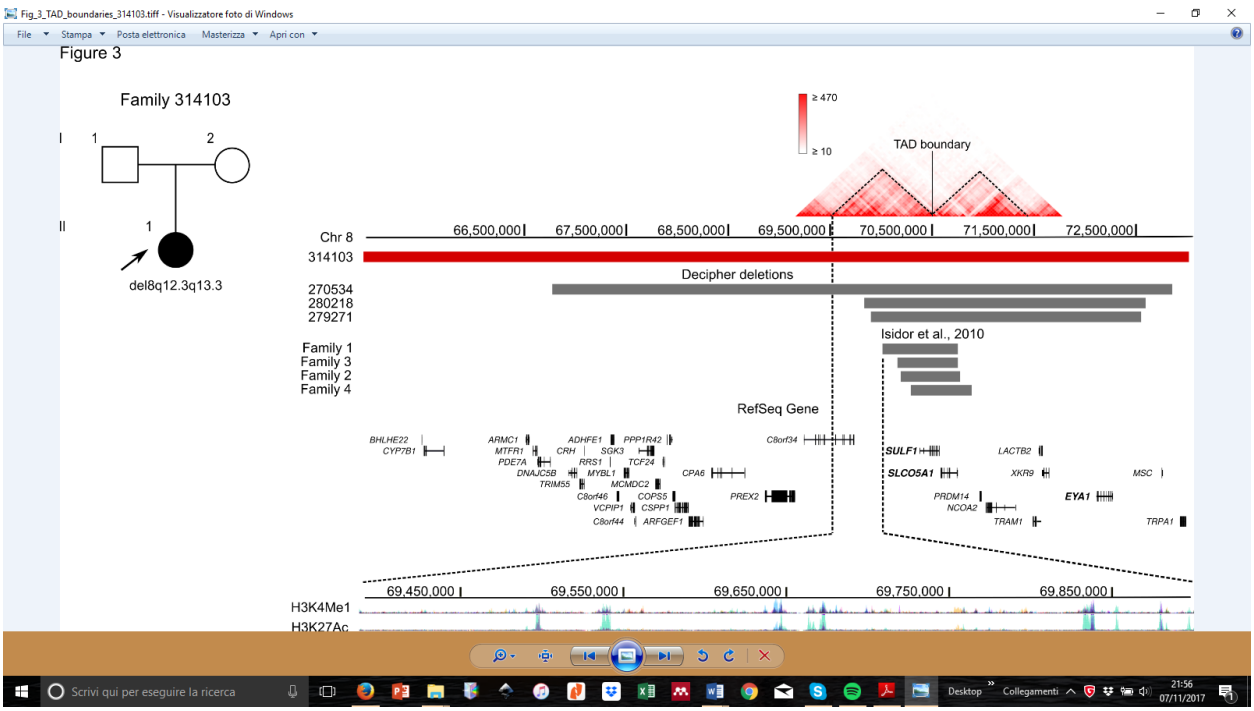


Fig_3_TAD_boundaries_314103.tiff - Visualizzatore foto di Windows
File Stampa Posta elettronica Masterizza Apri con
Figure 3
Family 314103
I 1 2
II 1
del8q12.3q13.3
≥ 470
≥ 10
TAD boundary
Chr 8
66,500,000 67,500,000 68,500,000 69,500,000 70,500,000 71,500,000 72,500,000
314103
Decipher deletions
270534
280218
279271
Isidor et al., 2010
Family 1
Family 3
Family 2
Family 4
RefSeq Gene
BHLHE22 CYP7B1 ARMC1 MTFR1 PDE7A DNAJC5B TRIM55 C8orf46 VCPIP1 C8orf44 ADHFE1 CRH RRS1 MYBL1 MCMDC2 COPS5 CSPP1 ARFGEF1 PPP1R42 SGK3 TCF24 CPA6 PREX2 C8orf34 SULF1 SLCO5A1 LACTB2 XKR9 MSC PRDM14 NCOA2 EYA1 TRAM1 TRPA1
69,450,000 69,550,000 69,650,000 69,750,000 69,850,000
H3K4Me1
H3K27Ac
Scrivi qui per eseguire la ricerca
Desktop Collegamenti 21:56 07/11/2017

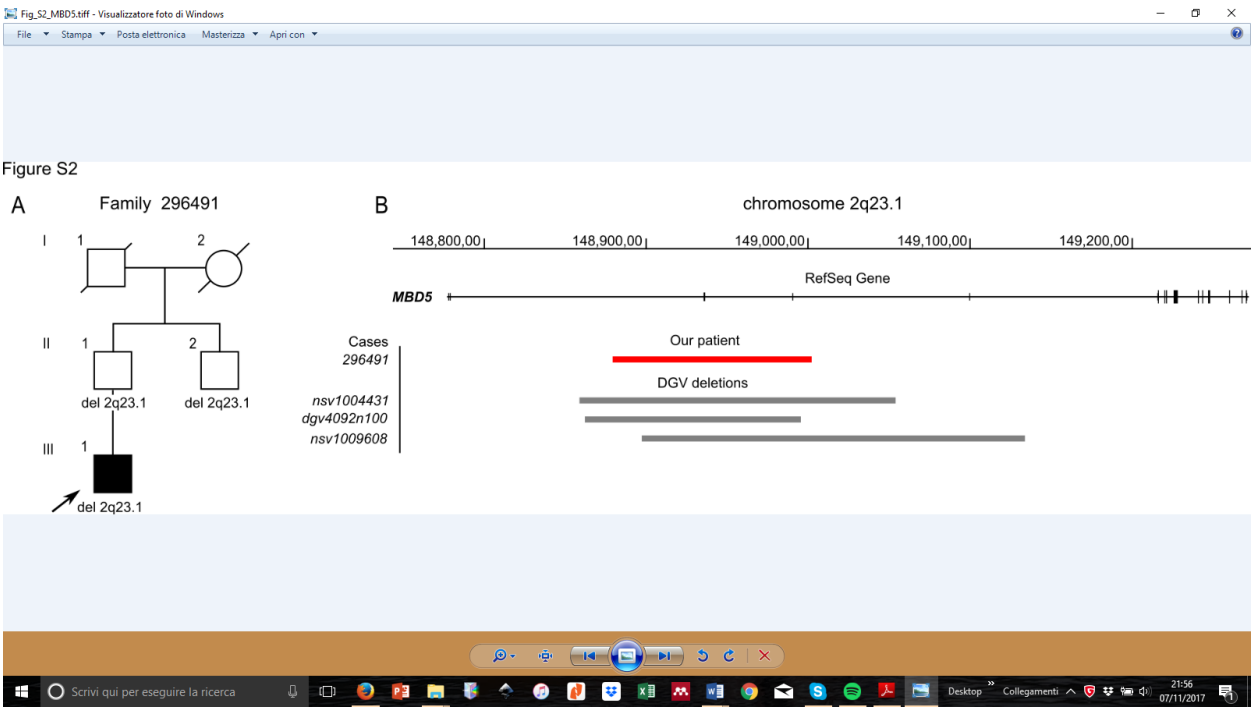

Fig_S2_MBD5.tiff - Visualizzatore foto di Windows
File
Stampa
Posta elettronica
Masterizza
Apri con

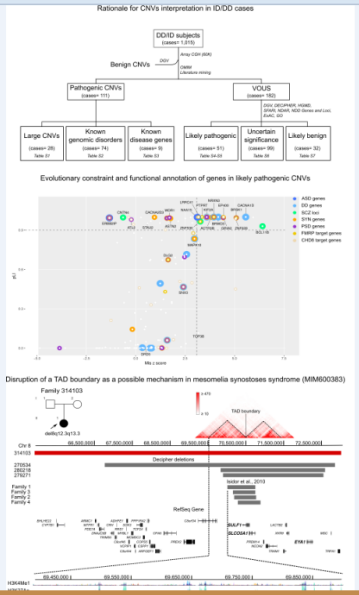
Figure S2

A
Family 296491
I
1
2
II
1
2
del 2q23.1
del 2q23.1
III
1
del 2q23.1
B
chromosome 2q23.1
148,800,00
148,900,00
149,000,00
149,100,00
149,200,00
RefSeq Gene
MBD5
Cases
296491
nsv1004431
dgv4092n100
nsv1009608
Our patient
DGV deletions

Scrivi qui per eseguire la ricerca
Desktop
Collegamenti
21:56
07/11/2017

Graphical_abstract.tiff - Visualizzatore foto di Windows

File Stampa Posta elettronica Masterizza Apri con



Scrivi qui per eseguire la ricerca

Desktop Collegamenti 21:56 07/11/2017