Whole exome sequencing is necessary to clarify ID/DD cases with de novo copy number variants of uncertain significance: Two proof-of-concept examples

Exome sequencing uncovers biallelic mutations in *TRAPPC9* and *VLDLR* and solve two syndromic intellectual disability cases with de novo CNVs.

Elisa Giorgio[1,*], Andrea Ciolfi[2,*], Elisa Biamino[3], Viviana Caputo[4], Eleonora Di Gregorio[5], Elga Fabia Belligni[3], Alessandro Calcia[1], Elena Gaidolfi[6], Alessandro Bruselles[2], Cecilia Mancini[1], Simona Cavalieri[5], Cristina Molinatto[3], Margherita Cirillo Silengo[3], Giovanni Battista Ferrero[3], Marco Tartaglia[2,7,§], Alfredo Brusco[1,5,§].

[1] University of Torino, Department of Medical Sciences, Turin, 10126, Italy

[2] Department of Hematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, 00161, Italy

[3] University of Torino, Department of Public Health and Pediatrics, Turin, 10126, Italy

[4] Department of Experimental Medicine, Sapienza University of Rome, Rome, 00161, Italy

[5] Città della Salute e della Scienza University Hospital, Medical Genetics Unit, Turin, 10126, Italy

[6] Centro Diagnostico Cernaia, Magnetic Resonance Unit, Turin, 10122, Italy

[7] Malattie Genetiche e Malattie rare, Ospedale Pediatrico Bambino Gesù IRCSS, Rome, 00146 Italy

*These authors equally contributed to this work

[§]These authors jointly directed this work

Corresponding author:
Alfredo Brusco, PhD,
Department of Medical Sciences,
University of Torino,
via Santena 19,
10126, Torino, Italy.
Phone: +390116334480;
Fax; +390116706582;
E-mail: alfredo.brusco@unito.it.

Keywords: whole exome sequencing, *de novo* CNV, intellectual disability, *VLDRL*, *TRAPPC9*.

ABSTRACT

We report on two sporadic cases with syndromic intellectual disability/developmental delay (ID/DD) carrying a *de novo* copy number variant (CNV): a 130-480 kb deletion spanning *ARHGAP12*, and a 200-345 kb duplication spanning the *CNOT6, SCGB3A1* and *FLT4* genes. Both rearrangements were considered variants of unknown significance (VOUS) although their *de novo* nature and the role of the encoded proteins suggested a possible clinical significance. Because of consanguinity in both families, we performed whole exome sequencing (WES), which allowed to identify a functionally relevant homozygous variant affecting a previously identified disease gene for rare syndromic ID/DD in each proband, *i.e.*, c.1423C>T (p.Arg377*) in the Trafficking Protein Particle Complex 9 (*TRAPPC9*), and c.154T>C (p.Cys52Arg) in the Very Low Density Lipoprotein Receptor (*VLDLR*). Four mutations affecting *TRAPPC9* had previously been reported, and the present finding further depicts this syndromic form of ID which includes microcephaly with brachycephaly, corpus callosum hypoplasia, facial features including round face, straight eyebrows, synophrys, deep set eyes, wide nasal bridge, and thin upper lip, and overweight. VLDLR-associated cerebellar hypoplasia (VLDLR-CH) is characterized by non-progressive congenital ataxia and moderate-to-profound intellectual disability. The c.154T>C (p.Cys52Arg) mutation was associated with a very mild form of ataxia, mild intellectual disability, cerebellar hypoplasia without cortical gyri simplification.

In conclusion, we report two novel cases with rare causes of autosomal recessive ID that document how the interpretation of *de novo* array-CGH variants represents a challenge in consanguineous families, where WES may become a mandatory diagnostic testing.

INTRODUCTION

Array-CGH is a widely used technology recommended as first-tier test for postnatal evaluation of individuals with intellectual disability/developmental delay (ID/DD), autism spectrum disorders (ASD), and/or multiple congenital anomalies (MCA) [Manning and others 2010; Miller and others 2010]. Pathogenic variants are detected in 15-20% of ID/DD patients [Vissers and others 2010b], who generally carry a deletion/duplication involving a known disease-associated genomic region or spanning one or more disease genes. Because the identification of unreported copy number variants (CNVs) raises challenges in their interpretation, the American College of Medical Genetics (ACMG) developed guidelines for their reporting [Kearney and others 2011]. Rearrangements should be listed as benign, pathogenic, or reported as variants of unknown clinical significance, this category being fairly broad and including findings later demonstrated to be either undoubtedly pathogenic or benign. Important recommendations to evaluate and clinically interpret a CNV include whether it comprises gene-rich regions or is void of genes as well as the type of genes involved. Of note, the *de novo* nature of a CNV has been considered an important indication of its involvement in neurodevelopmental and neuropsychiatric disorders [Levy and others 2011; Pinto and others 2010; Sanders and others 2011; Sebat and others 2007]. Other associations, including the higher prevalence of *de novo* variants reported in sporadic schizophrenia cases compared to controls (10% vs. 1.3%) [Xu and others 2012; Xu and others 2008], would support this interpretation.

Here, we report on two consanguineous families with probands exhibiting sporadic syndromic ID/DD for whom a *de novo* CNV had to be interpreted. In both cases, whole exome sequencing (WES) was crucial for a correct diagnosis, allowing to identify the disease-causing mutations, and reconsider each CNV as not the causative event underlying the disorder.

MATERIALS AND METHODS

Patients

In our survey of over 900 patients with ID/DD or multiple congenital anomalies referred for array-CGH diagnostic screening from 2008 to 2014, we identified two cases born to consanguineous parents having a *de novo* CNV. Patients performed diagnostic routine exams, which included a clinical genetic counseling. Both subjects executed magnetic resonance imaging disclosing unspecific abnormalities, while routine laboratory exams provided normal results. Karyotyping was performed on GTG-banded chromosomes from circulating leukocytes. Paternity was confirmed by microsatellite analyses using Profiler kit (Life Technologies). Patients were submitted to the Decipher database (ID codes 296553 and 296528; https://decipher.sanger.ac.uk). The study was performed with the approval of the Internal Review Board, and informed consents were obtained by patients' legal representatives.

Array-CGH analyses

Array-CGH was performed using a 60K whole-genome oligonucleotide microarray following the manufacturer's protocol (Agilent Technologies, Santa Clara, California, USA). Slides were scanned using a G2565BA scanner, and analyzed using Agilent CGH Analytics software v. 4.0.81 (Agilent Technologies Inc.) with the statistical algorithm ADM-2 and a sensitivity threshold of 6.0. Significant copy-number changes were identified by at least three consecutive aberrant probes. Reference human genomic DNA was GRCh37/hg19. Real-time PCR was used to confirm the array-CGH data and to further define the rearrangements (Supplemental fig. 1).

WES analysis

WES was outsourced at BGI-Shenzen using genomic DNA extracted from circulating leukocytes. Targeted enrichment was performed using Nimblegen SeqCap EZ Library v.3.0 (64 M) (Roche), and

captured libraries were loaded onto an Illumina HiSeq 2000 platform (Illumina). WES data analysis was performed using an in-house implemented pipeline [Cordeddu and others 2014; Kortüm and others 2015; Niceta and others 2015]. In brief, paired-end reads were aligned to human genome (UCSC GRCh37/hg19) with the Burrows–Wheeler Aligner (BWA V. 0.7.5a-r405) [Li and Durbin 2009], and presumed PCR duplicates were discarded using the Picard's MarkDuplicates utility (http://picard.sourceforge.net). The alignment process was refined by local realignment and base-quality-score recalibration steps by means of Genome Analysis Toolkit (GATK 3.2) [McKenna and others 2010]. GATK UnifiedGenotyper and HaplotypeCaller were used to identify single nucleotide polymorphisms (SNPs) and insertions/deletions (INDELs) [DePristo and others 2011]. Variants with quality score < 50 and quality-by-depth < 1.5 or resulting from 4 or more reads having ambiguous mapping (this number being greater than 10% of all aligned reads) were discarded. Remaining variants were then filtered against available public (dbSNP141, retaining only variants with MAF < 0.001 or with a known clinical association), and in-house databases (retaining variants with frequency < 5%). SnpEff toolbox v3.6 [Cingolani and others 2012] was used to predict the functional impact of variants, and retain missense/nonsense/frameshift changes, coding indels, and intronic variants at exon-intron junctions (within position -5/+5). Functional annotation of variants was performed by using snpEff v3.6 and dbNSFP2.8 [Cingolani and others 2012; Liu and others 2013].

Based on consanguinity, we assumed an autosomal recessive model of inheritance for both traits, and retained all the homozygous variants located within LoH genomic stretches using Homozygosity Mapper [Seelow and Schuelke 2012] (http://www.homozygositymapper.org/), setting 35 as the number of consecutive homozygous SNPs. Retained variants were prioritized according to their predicted functional impact (SVM radial score >0 or CADD score >15) [Kircher and others 2014; Liu and others 2013], and their biological and clinical relevance.

Sequence validation and segregation analyses were performed by Sanger sequencing using an ABI 3130XL and the ABI BigDye Terminator Sequencing Kit V.3.1 (Life Technologies). Sequences were examined using the SeqScape v2.6 Software (Life Technologies).

RESULTS

Clinics and neuroradiology

Patient 296553 was a 4 year-old girl born after an uneventful pregnancy. Parents were second degree cousins of Egyptian origin. She was referred to the pediatric genetics unit for severe developmental delay. At physical examination, she displayed microcephaly with brachycephaly (OFC 45 cm, < 3rd centile) and a peculiar facies characterized by round face, thin and horizontal eyebrows, synophrys, deep set eyes, wide nasal bridge and thin upper lip (Fig. 1A, B). She could walk with support; speech was absent and stereotypic movements were apparent (hand shaking, waving and body rocking). Brain magnetic resonance imaging (MRI) performed at 3 years showed severe corpus callosum thinning (Fig.1C) and a clear reduction of the white matter with poor myelination (Fig.1C-E); cerebellum was normal (Fig.1C, D). Independent walking was achieved at age of 5 years. At the age of 7 (last examination), the parents complained of frequent nocturnal awakenings and temper tantrums with self-injury; weight was 30 kg (97th centile), height 120 cm (50th centile), OFC 47 cm (< 3rd centile). She presented with severe ID, language limited to a few syllabi and motor stereotypies.

Patient 296528 was the second child of Moroccan origin first degree cousins. Family history was remarkable for a first degree cousin affected by severe ID (independent walking achieved at 10 years) and strabismus. Pregnancy was reported normal. She was born at 39 weeks of gestation with normal auxometric parameters (weight: 3,560 gr; length: 49 cm; OFC: 35 cm), APGAR scores were 9/9. Global developmental delay was diagnosed at the age of 2 years, when she achieved independent ambulation. At that time, neurological evaluation disclosed legs hypotonia and mild ataxia. She was

therefore referred for pediatric genetic evaluation: she displayed weight and length at 25th centile, ataxic wide-based ambulation, bilateral *pes planus*, difficulties in subtle manipulation; facial dysmorphism was not apparent (Fig.1F, G). Brain MRI detected severe cerebellar vermis hypoplasia with enlarged brain cerebrospinal fluid spaces. Cortical gyration was normal(Fig.1H-J). Further investigations, including electroencephalography, ophthalmological evaluation, and general and metabolic workup (blood count, CPK, lipid profile, serum albumin, liver enzymes, transferrin, lactate, plasma acylcarnitine, transferrin isoelectrofocusing, and VitE) did not provide informative data for diagnosis. At the age of 6 years (last evaluation), height was 107 cm (10th centile), OFC 50 cm (25th centile); gait ataxia was regressed, and the patient walked independently without aid. A mild dysmetria was present at the finger-nose and heel-shin tests. Dysarthria was present. Ophthalmological exam was normal.

Array-CGH

Array-CGH analysis documented a *de novo* 134-483 kb deletion on 10p11.22 in case 296553 [arr 10p11.22(31,817,746x2,32,095,083-32,229,198x1,32,300,151), hg19] spanning the *ARHGAP12* gene (MIM 610577), and a *de novo* 200-345 kb duplication on 5q35.3 in case 296528 [arr 5q35.3(179,807,078x2, 179,878,423-180,075,503x3,180,152,402x2), hg19] encompassing the *CNOT6* (MIM 608951), *SCGB3A1* (MIM 606500) and *FLT4* (MIM 136352) genes (Fig. 2 and supplemental fig.1, 2). Real-time PCR assays confirmed the rearrangements and their *de novo* origin, although we did not further define the limits of the duplicated genomic region in case 296528 (Supplemental fig.1). Decipher database reports three cases with a deletion and three with a duplication spanning *ARHGAP12*; all records referred to large rearrangements (3.5-10 Mb) encompassing multiple genes (Supplemental Fig.2). Several rearrangements spanning *CNOT6, SCGB3A1* and *FLT4* are reported in

Decipher database, but all are large (>10 Mb) suggesting many genes may contribute to those phenotypes.

Exome sequencing

WES statistics are reported in Supplemental table 1. Data annotation predicted 12,859 (case 296553) and 12,476 (case 296528) high-quality variants having functional impact (*i.e.*, non-synonymous and splice site changes). Among them, 2,353 and 2,134 private, rare (minor allele frequency < 0.001) or clinically associated changes were further analyzed. Variants were filtered to retain rare or private homozygous sequence changes located within LoH regions, and *in silico* analyses of the predicted functional impact of individual variants and biological relevance of the encoded proteins allowed to identify an excellent disease gene candidate in each patient (Supplemental table 2 and Supplemental Fig. 2). A nonsense change, c.1423C>T (p.Arg377*) (rs267607136, flagged as clinically associated), was identified in *Trafficking Protein Particle Complex 9* (*TRAPPC9*, MIM 611966) in case 296553 (Fig.2). *TRAPPC9* encodes a protein implicated in NF-kB activation, and five inactivating mutations in this gene have been reported to underlie a rare, recessive non-syndromic ID associated with microcephaly, mild cerebral white matter hypoplasia, and corpus callosum hypoplasia (MIM 613192) (Fig. 3), which fitted well with the clinical features exhibited by the proband.

Case 296528 was homozygous for a missense change, c.154T>C (p.Cys52Arg), in the *Very Low Density Lipoprotein Receptor* gene (*VLDLR*, MIM 192977) (Fig. 2). The affected residue is highly conserved (Supplemental Fig. 3), involved in a intramolecular disulfide bridge required for proper receptor function, and resides in the ligand-binding type repeat (LBTR) region. Consistently, the substitution was predicted to be deleterious. Homozygous or compound heterozygous mutations in *VLDLR* have been reported to cause cerebellar ataxia, mental retardation and disequilibrium syndrome

type 1 (CAMRQ1; MIM 224050) (Fig. 3), a disorder with features that overlap those of our patient. In both probands, Sanger sequencing validated both sequence changes and segregation.

DISCUSSION

Guidelines for investigating causality of unannotated CNVs take in consideration their *de novo* origin among the most important factors [Buysse and others 2009; Gijsbers and others 2009; Gijsbers and others 2011; Koolen and others 2009; Lee and others 2007; Miller and others 2010]. Here we report on two cases in whom array-CGH identified CNVs that were initially suspected to be causative of the disease because of their *de novo* occurrence in each proband. In the first case, a heterozygous deletion encompassed *ARGAPH12*, which codes for a Rho-GTPase-activating protein negatively controlling function of Rho subfamily members. Rho-GTPases have been identified as key regulators of cytoskeleton structural changes in many cell types, including neurons [Heasman and Ridley 2008], and play a major role in dendritic spine development [Tolias and others 2011]. In analogy to other proteins of the same family involved in ID (e.g., oligophrenin) and playing important roles in the developing axons and growth cones, *ARHGAP12* haploinsufficiency was originally hypothesized to have a causative role in the disease. In the second case, the duplicated region encompassed three genes: *FLT4* encodes a tyrosine kinase receptor for vascular endothelial growth factors C and D that is apparently involved in lymphangiogenesis and maintenance of the lymphatic endothelium. Mutations in this gene cause autosomal dominant lymphedema type IA (MIM 153100) due to a loss of function/dominant negative mechanism [Connell and others 2009; Ferrell and others 1998; Gordon and others 2013]. Our patient did not show any sign of lymphedema or lymphatic system involvement (e.g., pleural effusions, intestinal lymphangiectasia, ascites). Lymphoscintigraphy was not appropriate due to unjustified invasiveness. We did not notice dysplastic nails, anomalous palm-plantar creases or any obvious venous malformation. These findings support the idea that the duplication of *FLT4* is not associated

with a pathogenic phenotype. No Mendelian disease has been associated with *SCGB3A*, which encodes a secretoglobin [Krop and others 2001]. The *CNOT6* gene encodes a subunit of the Carbon Catabolite Repressor Protein 4 (CCR4-NOT) core transcriptional regulation complex. CCR4a is implicated in cell proliferation, cell cycle arrest and senescence, and it is required for foci formation [Chen and others 2011; Chen and others 2002]. Given the role of transcription regulation in the pathogenesis of ID/DD [van Bokhoven 2011], and the widespread expression of *CNOT6*, we originally considered its duplication as possibly causative for the condition, even if classified as a variant of unknown significance.

Recent publications show that small *de novo* imbalances must not automatically be classified as likely casual for the investigated phenotype in the absence of strong evidence from other data sources, and rearrangements below 500 kb have to be considered carefully. An historical example of *de novo* CNV wrongly assigned as pathogenic is presented by the 250 kb deletion in *MACROD2*, which was described in a patient with Kabuki syndrome, later found to be mutated in the *MLL2* gene [Maas and others 2007; Paulussen and others 2011]. More recently, a *de novo* 86.5 kb deletion was reported pathogenic in a patient with ID and eye disorder, because it harbored *AMBRA1*, a gene expressed in the neural retina and brain [Fimia and others 2007]. Subsequent accurate clinical evaluation of the patient suggested a possible diagnosis within the clinical spectrum of CHARGE syndrome, which was confirmed by the identification of the disease causative mutation in *CHD7* [Vissers and others 2004].

Our cases further support the caveats concerning small *de novo* CNVs. This concern particularly applies to ID/DD-associated traits described in the context of consanguinity. In these cases, the analysis of the exome, particularly when restricted to the scanning of genes that have been associated with Mendelian disorders (*i.e.*, clinical exome), is particularly informative. Here, we document that WES analysis allowed to identify the causal molecular lesion in both cases. In the first family of Egyptian origin, a homozygous nonsense mutation (c.1423C>T; p.Arg377*) in *TRAPPC9* was recognized.

TRAPPC9 has been implicated in NF-kB activation, and it is possibly involved in intracellular trafficking. The same truncating lesion had previously been reported in families from Pakistan, Syria and of Arab-Israeli origin [Abou Jamra and others 2011; Mir and others 2009; Mochida and others 2009]. Only five mutations are known in this gene, all with a predicted inactivating effect (Fig. 3). The *TRAPPC9* mutation-associated phenotype was initially reported as non-syndromic ID with postnatal microcephaly [Mir and others 2009; Mochida and others 2009; Philippe and others 2009]. However, consistent with the present findings, more recent reports provided evidence that loss of TRAPPC9 function underlies a syndromic form of ID with distinctive facial features (brachycephaly, round face, straight eyebrows, synophrys, deep set eyes, wide nasal bridge, and thin upper lip), true or relative microcephaly, MRI brain anomalies (corpus callosum hypoplasia, reduced white matter volume with multifocal hyperintensity), and overweight [Marangi and others 2013]. Frequent sleep awakenings and motor stereotypies, represent also variably occurring features [Abou Jamra and others 2011; Marangi and others 2013].

In the second family, we identified a homozygous previously unreported missense change, c.154T>C (p.Cys52Arg) in the *VLDLR* gene. In analogy with Low Density Lipoprotein Receptor (LDLR), the binding domain of VLDLR to lipoproteins contains seven tandem repeated cysteine rich domains at its aminoterminus [Fass and others 1997](Fig. 3). Each repeat of ∼40 amino acids, contains two loops stabilized by three disulphide bridges which are required for proper folding of the domain. Cys^{52} is predicted to be involved in an intramolecular disulfide bond with Cys^{67} (http://www.uniprot.org/uniprot/P98155), and loss of this disulfide bridge is expected to result in protein misfolding and its degradation by the ER-associated protein degradation machinery (ERAD) [Ali and others 2012]. Eleven mutations in this gene have been reported most with a predicted loss of function mechanism (Fig. 3). Only three missense changes are known, all apparently associated with a classical CAMRQ1 phenotype. The clinical phenotype associated with *VLDRL* mutations is relatively

homogeneous and includes non-progressive truncal ataxia, dysarthria, moderate to profound intellectual disability, and *pes planus*. MRI shows cerebellar hypoplasia (mainly vermian) and a simplification of cortical gyri. Other symptoms, such as epilepsy, are variably associated. Some mutation has been associated with quadrupedal locomotion [Ozcelik and others 2008; Tan 2006; Turkmen and others 2008] although this was suggested to be a physical adaptation [Sonmez and others 2013]. Of note, our patient exhibited a milder phenotype, which may be specifically associated with the type and location of mutation which that might result in a receptor with not completely impaired function (see Fig.3). Notably, MRI showed hypoplasia of cerebellar vermis, but cerebral gyration was normal, in contrast with all reported cases.

In conclusion, diagnosis in both patients would have been missed or mislead, based on the array-CGH data interpretation. This report further emphasizes the utility of WES to explore the possible occurrence of rare genetic disorders in consanguineous families even if *de novo* CNVs are found. To avoid misinterpretations, WES should be used together with array-CGH as a first-tier diagnostic tool in consanguineous cases [Vissers and others 2010a].

COMPETING INTERESTS

Dr. Elena Gaidolfi is an employee of the Centro Diagnostico Cernaia, a private diagnostic center.

ACKNOWLEDGMENTS

We are grateful to all family members who contributed to the study. This work was funded by MURST 60% (to A. Brusco), Istituto Superiore di Sanità (ricerca corrente 2013 to M.T.) and the financial support from the company BVLGARI. We thank CINECA for computational resources (WES data analysis). This study makes use of data generated by the DECIPHER Consortium. A full list of centers

who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by the Wellcome Trust.

FIGURE LEGENDS

Figure 1. Clinical features of the two affected subjects included in the study. Proband 296553 (upper panels) exhibits features and signs described *TRAPPC9* mutation-associated subjects, including round face, brachycephaly, thin and horizontal eyebrows, synophrys, deep set eyes, thin upper lip (A, B). Brain magnetic resonance imaging (MRI) performed at 3 years showed severe corpus callosum thinning (C, T1-weighted sagittal section, asterisk) and a clear reduction of the white matter with poor myelination (C-E); cerebellum was apparently unaffected (D T2 weighted, E T1-weighted, coronal and axial sections). Proband 296528 (bottom panels) did not show facial dysmorphisms (F and G). Brain MRI detected severe cerebellar vermis hypoplasia (H, T1-weighted sagittal section; asterisk) with enlarged liquoral spaces and IV ventricle (H, flair coronal section, asterisk). Cortical gyration was unaffected (J, T1-weighted).

Figure 2. Genealogical trees and molecular data. Family trees of the two consanguineous families (A, D) are shown together with the array-CGH results (B, E). Sequence chromatograms showing the disease-causing mutations, c.1423C>T (p.Arg377*) in the *TRAPPC9* gene and c.154T>C (p.Cys52Arg) in the *VLDRL* gene, are reported in panels C and F, respectively.

Figure 3. Mutational spectrum of *TRAPPC9* and *VLDRL* genes. *TRAPPC9* (upper panel) and *VLDLR* (bottom panel) gene and protein structures are shown. Black boxes represent coding exons and untranslated exons (smaller boxes). Point mutations described in the literature are reported color coded by type (see legend). Mutations described in this paper are boxed. All mutations have been reported to occur as homozygous changes, with the exception of the c.1459G>T (p.D521H) and c.1711dupT (p.Y571LfsX7) in *VLDRL* that were documented in a compound heterozygous case. VLDLR functional domains are reported: LDLa, LDL-receptor class A; EGFCA, epidermal growth factor Calcium-

binding-like domain (EGFCA); LY, low-density lipoprotein-receptor YWTD domain; EGF, epidermal growth factor domain; TM, transmembrane domain.

Supplemental figure 1. Real-time PCR analysis performed to confirm the array-CGH results. For each patient, array-CGH data are reported with the red and blue bars indicating the minimal deleted and minimal duplicated regions, respectively. Flanking green bars represent regions with normal array-CGH signals. Genes spanning the rearrangement are shown with black (within the rearrangement) or grey (outside the rearrangement) arrows. The position of real-time PCR assays (UPL probe assay, Roche Diagnostics, Mannheim, Germany) used to validate the array-CGH data are represented by vertical red bars. Histograms show the result for each assay (see flanking table for conditions). In both cases, real-time PCR documented that the rearrangement was *de novo*. In case 296553, the deletion involved the entire *ARHGAP12* gene. In case 296528, the uncertainty in the duplication definition did not allow to establish if the upstream region of the *FLT4* gene was included.

Supplemental figure 2. Decipher cases with overlapping genomic rearrangements. The rearranged genomic regions in patients 296553 and 296528 is enlarged in panels A and B. Below, we report Decipher database cases with overlapping rearrangements (red, deletions; blue, duplications).

Supplemental Figure 3. Homozygosity mapping analysis. Plot of homozygosity regions (red bars) identified in patients 296553 (A) and 296528 (B) using HomozygosityMapper tool. The two disease causative variants identified in each patient are localized in long regions of homozygosity spanning about 10 Mb (*TRAPPC9*) and 4.8 Mb (*VLDLR*).

Supplemental Figure 4. Multiple sequence alignment of *VLDLR* orthologues showing conservation of Cys^{52}. The amino acid stretch encompassing the affected residue is shown (residues 33 to 82, in the human VLDLR protein).

REFERENCES

Abou Jamra R, Wohlfart S, Zweier M, Uebe S, Priebe L, Ekici A, Giesebrecht S, Abboud A, Al Khateeb MA, Fakher M, Hamdan S, Ismael A, Muhammad S, Nothen MM, Schumacher J, Reis A. 2011. Homozygosity mapping in 64 Syrian consanguineous families with non-specific intellectual disability reveals 11 novel loci and high heterogeneity. Eur J Hum Genet 19(11):1161-1166.

Ali BR, Silhavy JL, Gleeson MJ, Gleeson JG, Al-Gazali L. 2012. A missense founder mutation in VLDLR is associated with Dysequilibrium Syndrome without quadrupedal locomotion. BMC medical genetics 13:80.

Buysse K, Delle Chiaie B, Van Coster R, Loeys B, De Paepe A, Mortier G, Speleman F, Menten B. 2009. Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. Eur J Med Genet 52(6):398-403.

Chen C, Ito K, Takahashi A, Wang G, Suzuki T, Nakazawa T, Yamamoto T, Yokoyama K. 2011. Distinct expression patterns of the subunits of the CCR4-NOT deadenylase complex during neural development. Biochemical and biophysical research communications 411(2):360-364.

Chen J, Chiang YC, Denis CL. 2002. CCR4, a 3'-5' poly(A) RNA and ssDNA exonuclease, is the catalytic component of the cytoplasmic deadenylase. EMBO J 21(6):1414-1426.

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6(2):80-92.

Connell FC, Ostergaard P, Carver C, Brice G, Williams N, Mansour S, Mortimer PS, Jeffery S, Lymphoedema C. 2009. Analysis of the coding regions of VEGFR3 and VEGFC in Milroy disease and other primary lymphoedemas. Human genetics 124(6):625-631.

Cordeddu V, Redeker B, Stellacci E, Jongejan A, Fragale A, Bradley TE, Anselmi M, Ciolfi A, Cecchetti S, Muto V, Bernardini L, Azage M, Carvalho DR, Espay AJ, Male A, Molin AM, Posmyk R, Battisti C, Casertano A, Melis D, van Kampen A, Baas F, Mannens MM, Bocchinfuso G, Stella L, Tartaglia M, Hennekam RC. 2014. Mutations in ZBTB20 cause Primrose syndrome. Nat Genet 46(8):815-817.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43(5):491-498.

Fass D, Blacklow S, Kim PS, Berger JM. 1997. Molecular basis of familial hypercholesterolaemia from structure of LDL receptor module. Nature 388(6643):691-693.

Ferrell RE, Levinson KL, Esman JH, Kimak MA, Lawrence EC, Barmada MM, Finegold DN. 1998. Hereditary lymphedema: evidence for linkage and genetic heterogeneity. Hum Mol Genet 7(13):2073-2078.

Fimia GM, Stoykova A, Romagnoli A, Giunta L, Di Bartolomeo S, Nardacci R, Corazzari M, Fuoco C, Ucar A, Schwartz P, Gruss P, Piacentini M, Chowdhury K, Cecconi F. 2007. Ambra1 regulates autophagy and development of the nervous system. Nature 447(7148):1121-1125.

Gijsbers AC, Lew JY, Bosch CA, Schuurs-Hoeijmakers JH, van Haeringen A, den Hollander NS, Kant SG, Bijlsma EK, Breuning MH, Bakker E, Ruivenkamp CA. 2009. A new diagnostic workflow for patients with mental retardation and/or multiple congenital abnormalities: test arrays first. Eur J Hum Genet 17(11):1394-1402.

Gijsbers AC, Schoumans J, Ruivenkamp CA. 2011. Interpretation of array comparative genome hybridization data: a major challenge. Cytogenet Genome Res 135(3-4):222-227.

Gordon K, Spiden SL, Connell FC, Brice G, Cottrell S, Short J, Taylor R, Jeffery S, Mortimer PS, Mansour S, Ostergaard P. 2013. FLT4/VEGFR3 and Milroy disease: novel mutations, a review of published variants and database update. Human mutation 34(1):23-31.

385 Heasman SJ, Ridley AJ. 2008. Mammalian Rho GTPases: new insights into their functions from in vivo studies.
386 Nat Rev Mol Cell Biol 9(9):690-701.
387 Kearney HM, South ST, Wolff DJ, Lamb A, Hamosh A, Rao KW, Working Group of the American College of
388 Medical G. 2011. American College of Medical Genetics recommendations for the design and
389 performance expectations for clinical genomic copy number microarrays intended for use in the
390 postnatal setting for detection of constitutional abnormalities. Genetics in medicine : official journal of
391 the American College of Medical Genetics 13(7):676-679.
392 Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. 2014. A general framework for estimating the
393 relative pathogenicity of human genetic variants. Nat Genet 46(3):310-315.
394 Koolen DA, Pfundt R, de Leeuw N, Hehir-Kwa JY, Nillesen WM, Neefs I, Scheltinga I, Sistermans E, Smeets D,
395 Brunner HG, van Kessel AG, Veltman JA, de Vries BB. 2009. Genomic microarrays in mental retardation:
396 a practical workflow for diagnostic applications. Human mutation 30(3):283-292.
397 Kortüm F, Caputo V, Bauer CK, Stella S, Ciolfi A, Alawi M, Bocchinfuso G, Flex E, Paolacci S, Dentici ML,
398 Grammatico P, Korenke GC, Leuzzi V, Mowat D, Nair LDV, Nguyen TTN, Thierry P, White SM,
399 Dallapiccola B, Pizzuti A, Campeau PM, Tartaglia M, Kutsche K. 2015. Mutations in KCNH1 and
400 ATP6V1B2 cause Zimmermann-Laband syndrome. Nature Genetics in press.
401 Krop IE, Sgroi D, Porter DA, Lunetta KL, LeVangie R, Seth P, Kaelin CM, Rhei E, Bosenberg M, Schnitt S, Marks JR,
402 Pagon Z, Belina D, Razumovic J, Polyak K. 2001. HIN-1, a putative cytokine highly expressed in normal
403 but not cancerous mammary epithelial cells. Proc Natl Acad Sci U S A 98(17):9796-9801.
404 Lee C, Iafrate AJ, Brothman AR. 2007. Copy number variations and clinical cytogenetic diagnosis of
405 constitutional disorders. Nat Genet 39(7 Suppl):S48-54.
406 Levy D, Ronemus M, Yamrom B, Lee YH, Leotta A, Kendall J, Marks S, Lakshmi B, Pai D, Ye K, Buja A, Krieger A,
407 Yoon S, Troge J, Rodgers L, Iossifov I, Wigler M. 2011. Rare de novo and transmitted copy-number
408 variation in autistic spectrum disorders. Neuron 70(5):886-897.
409 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics
410 25(14):1754-1760.
411 Liu X, Jian X, Boerwinkle E. 2013. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional
412 predictions and annotations. Human mutation 34(9):E2393-2402.
413 Maas NM, Van de Putte T, Melotte C, Francis A, Schrander-Stumpel CT, Sanlaville D, Genevieve D, Lyonnet S,
414 Dimitrov B, Devriendt K, Fryns JP, Vermeesch JR. 2007. The C20orf133 gene is disrupted in a patient
415 with Kabuki syndrome. J Med Genet 44(9):562-569.
416 Manning M, Hudgins L, Professional P, Guidelines C. 2010. Array-based technology and recommendations for
417 utilization in medical genetics practice for detection of chromosomal abnormalities. Genetics in
418 medicine : official journal of the American College of Medical Genetics 12(11):742-745.
419 Marangi G, Leuzzi V, Manti F, Lattante S, Orteschi D, Pecile V, Neri G, Zollino M. 2013. TRAPPC9-related
420 autosomal recessive intellectual disability: report of a new mutation and clinical phenotype. Eur J Hum
421 Genet 21(2):229-232.
422 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly
423 M, DePristo MA. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-
424 generation DNA sequencing data. Genome Res 20(9):1297-1303.
425 Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, Church DM, Crolla JA, Eichler EE,
426 Epstein CJ, Faucett WA, Feuk L, Friedman JM, Hamosh A, Jackson L, Kaminsky EB, Kok K, Krantz ID, Kuhn
427 RM, Lee C, Ostell JM, Rosenberg C, Scherer SW, Spinner NB, Stavropoulos DJ, Tepperberg JH, Thorland
428 EC, Vermeesch JR, Waggoner DJ, Watson MS, Martin CL, Ledbetter DH. 2010. Consensus statement:
429 chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental
430 disabilities or congenital anomalies. Am J Hum Genet 86(5):749-764.
431 Mir A, Kaufman L, Noor A, Motazacker MM, Jamil T, Azam M, Kahrizi K, Rafiq MA, Weksberg R, Nasr T, Naeem
432 F, Tzschach A, Kuss AW, Ishak GE, Doherty D, Ropers HH, Barkovich AJ, Najmabadi H, Ayub M, Vincent

JB. 2009. Identification of mutations in TRAPPC9, which encodes the NIK- and IKK-beta-binding protein, in nonsyndromic autosomal-recessive mental retardation. Am J Hum Genet 85(6):909-915.

Mochida GH, Mahajnah M, Hill AD, Basel-Vanagaite L, Gleason D, Hill RS, Bodell A, Crosier M, Straussberg R, Walsh CA. 2009. A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. Am J Hum Genet 85(6):897-902.

Niceta M, Stellacci E, Gripp KW, Zampino G, Kousi M, Anselmi M, Traversa A, Ciolfi A, Stabley D, Bruselles A, Caputo V, Cecchetti S, Prudente S, Fiorenza MT, Boitani C, Philip N, Niyazov D, Leoni C, Nakane T, Keppler-Noreuil K, Braddock SR, Gillessen-Kaesbach G, Palleschi A, Campeau PM, Lee BHL, Pouponnot C, Stella L, Bocchinfuso G, Katsanis N, Sol-Church K, Tartaglia M. 2015. Mutations impairing GSK3-mediated MAF phosphorylation cause cataract, deafness, intellectual disability, seizures, and a Down syndrome-like facies. . American Journal of Human Genetics in press.

Ozcelik T, Akarsu N, Uz E, Caglayan S, Gulsuner S, Onat OE, Tan M, Tan U. 2008. Mutations in the very low-density lipoprotein receptor VLDLR cause cerebellar hypoplasia and quadrupedal locomotion in humans. Proc Natl Acad Sci U S A 105(11):4232-4236.

Paulussen AD, Stegmann AP, Blok MJ, Tserpelis D, Posma-Velter C, Detisch Y, Smeets EE, Wagemans A, Schrander JJ, van den Boogaard MJ, van der Smagt J, van Haeringen A, Stolte-Dijkstra I, Kerstjens-Frederikse WS, Mancini GM, Wessels MW, Hennekam RC, Vreeburg M, Geraedts J, de Ravel T, Fryns JP, Smeets HJ, Devriendt K, Schrander-Stumpel CT. 2011. MLL2 mutation spectrum in 45 patients with Kabuki syndrome. Human mutation 32(2):E2018-2025.

Philippe O, Rio M, Carioux A, Plaza JM, Guigue P, Molinari F, Boddaert N, Bole-Feysot C, Nitschke P, Smahi A, Munnich A, Colleaux L. 2009. Combination of linkage mapping and microarray-expression analysis identifies NF-kappaB signaling defect as a cause of autosomal-recessive mental retardation. Am J Hum Genet 85(6):903-908.

Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Almeida J, Bacchelli E, Bader GD, Bailey AJ, Baird G, Battaglia A, Berney T, Bolshakova N, Bolte S, Bolton PF, Bourgeron T, Brennan S, Brian J, Bryson SE, Carson AR, Casallo G, Casey J, Chung BH, Cochrane L, Corsello C, Crawford EL, Crossett A, Cytrynbaum C, Dawson G, de Jonge M, Delorme R, Drmic I, Duketis E, Duque F, Estes A, Farrar P, Fernandez BA, Folstein SE, Fombonne E, Freitag CM, Gilbert J, Gillberg C, Glessner JT, Goldberg J, Green A, Green J, Guter SJ, Hakonarson H, Heron EA, Hill M, Holt R, Howe JL, Hughes G, Hus V, Igliozzi R, Kim C, Klauck SM, Kolevzon A, Korvatska O, Kustanovich V, Lajonchere CM, Lamb JA, Laskawiec M, Leboyer M, Le Couteur A, Leventhal BL, Lionel AC, Liu XQ, Lord C, Lotspeich L, Lund SC, Maestrini E, Mahoney W, Mantoulan C, Marshall CR, McConachie H, McDougle CJ, McGrath J, McMahon WM, Merikangas A, Migita O, Minshew NJ, Mirza GK, Munson J, Nelson SF, Noakes C, Noor A, Nygren G, Oliveira G, Papanikolaou K, Parr JR, Parrini B, Paton T, Pickles A, Pilorge M, Piven J, Ponting CP, Posey DJ, Poustka A, Poustka F, Prasad A, Ragoussis J, Renshaw K, Rickaby J, Roberts W, Roeder K, Roge B, Rutter ML, Bierut LJ, Rice JP, Salt J, Sansom K, Sato D, Segurado R, Sequeira AF, Senman L, Shah N, Sheffield VC, Soorya L, Sousa I, Stein O, Sykes N, Stoppioni V, Strawbridge C, Tancredi R, Tansey K, Thiruvahindrapduram B, Thompson AP, Thomson S, Tryfon A, Tsiantis J, Van Engeland H, Vincent JB, Volkmar F, Wallace S, Wang K, Wang Z, Wassink TH, Webber C, Weksberg R, Wing K, Wittemeyer K, Wood S, Wu J, Yaspan BL, Zurawiecki D, Zwaigenbaum L, Buxbaum JD, Cantor RM, Cook EH, Coon H, Cuccaro ML, Devlin B, Ennis S, Gallagher L, Geschwind DH, Gill M, Haines JL, Hallmayer J, Miller J, Monaco AP, Nurnberger JI, Jr., Paterson AD, Pericak-Vance MA, Schellenberg GD, Szatmari P, Vicente AM, Vieland VJ, Wijsman EM, Scherer SW, Sutcliffe JS, Betancur C. 2010. Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466(7304):368-372.

Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, Chu SH, Moreau MP, Gupta AR, Thomson SA, Mason CE, Bilguvar K, Celestino-Soper PB, Choi M, Crawford EL, Davis L, Wright NR, Dhodapkar RM, DiCola M, DiLullo NM, Fernandez TV, Fielding-Singh V, Fishman DO, Frahm S, Garagaloyan R, Goh GS, Kammela S, Klei L, Lowe JK, Lund SC, McGrew AD, Meyer KA, Moffat WJ,

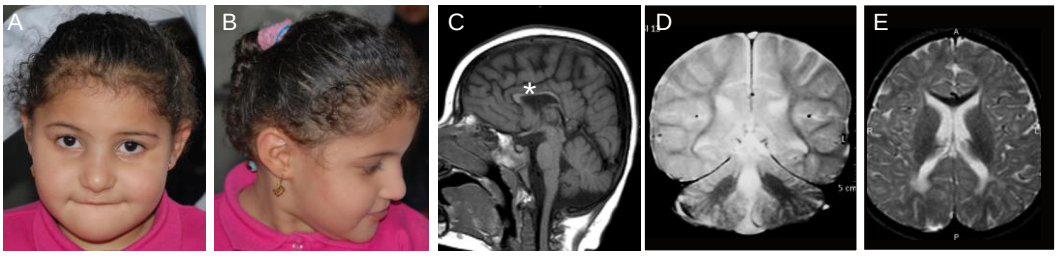
481 Murdoch JD, O'Roak BJ, Ober GT, Pottenger RS, Raubeson MJ, Song Y, Wang Q, Yaspan BL, Yu TW,
482 Yurkiewicz IR, Beaudet AL, Cantor RM, Curland M, Grice DE, Gunel M, Lifton RP, Mane SM, Martin DM,
483 Shaw CA, Sheldon M, Tischfield JA, Walsh CA, Morrow EM, Ledbetter DH, Fombonne E, Lord C, Martin
484 CL, Brooks AI, Sutcliffe JS, Cook EH, Jr., Geschwind D, Roeder K, Devlin B, State MW. 2011. Multiple
485 recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly
486 associated with autism. Neuron 70(5):863-885.
487 Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta
488 A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimaki T, Ledbetter D, Gregersen PK,
489 Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam
490 TC, Ye K, Wigler M. 2007. Strong association of de novo copy number mutations with autism. Science
491 316(5823):445-449.
492 Seelow D, Schuelke M. 2012. HomozygosityMapper2012--bridging the gap between homozygosity mapping
493 and deep sequencing. Nucleic Acids Res 40(Web Server issue):W516-520.
494 Sonmez FM, Gleeson JG, Celep F, Kul S. 2013. The very low density lipoprotein receptor-associated
495 pontocerebellar hypoplasia and dysmorphic features in three Turkish patients. J Child Neurol
496 28(3):379-383.
497 Tan U. 2006. A new syndrome with quadrupedal gait, primitive speech, and severe mental retardation as a live
498 model for human evolution. Int J Neurosci 116(3):361-369.
499 Tolias KF, Duman JG, Um K. 2011. Control of synapse development and plasticity by Rho GTPase regulatory
500 proteins. Progress in neurobiology 94(2):133-148.
501 Turkmen S, Hoffmann K, Demirhan O, Aruoba D, Humphrey N, Mundlos S. 2008. Cerebellar hypoplasia, with
502 quadrupedal locomotion, caused by mutations in the very low-density lipoprotein receptor gene. Eur J
503 Hum Genet 16(9):1070-1074.
504 van Bokhoven H. 2011. Genetic and epigenetic networks in intellectual disabilities. Annual review of genetics
505 45:81-104.
506 Vissers LE, de Ligt J, Gilissen C, Janssen I, Steehouwer M, de Vries P, van Lier B, Arts P, Wieskamp N, del Rosario
507 M, van Bon BW, Hoischen A, de Vries BB, Brunner HG, Veltman JA. 2010a. A de novo paradigm for
508 mental retardation. Nat Genet 42(12):1109-1112.
509 Vissers LE, de Vries BB, Veltman JA. 2010b. Genomic microarrays in mental retardation: from copy number
510 variation to gene, from research to diagnosis. J Med Genet 47(5):289-297.
511 Vissers LE, van Ravenswaaij CM, Admiraal R, Hurst JA, de Vries BB, Janssen IM, van der Vliet WA, Huys EH, de
512 Jong PJ, Hamel BC, Schoenmakers EF, Brunner HG, Veltman JA, van Kessel AG. 2004. Mutations in a
513 new member of the chromodomain gene family cause CHARGE syndrome. Nat Genet 36(9):955-957.
514 Xu B, Ionita-Laza I, Roos JL, Boone B, Woodrick S, Sun Y, Levy S, Gogos JA, Karayiorgou M. 2012. De novo gene
515 mutations highlight patterns of genetic and neural complexity in schizophrenia. Nat Genet 44(12):1365-
516 1369.
517 Xu B, Roos JL, Levy S, van Rensburg EJ, Gogos JA, Karayiorgou M. 2008. Strong association of de novo copy
518 number mutations with sporadic schizophrenia. Nat Genet 40(7):880-885.

519

520

Figure 1

Patient 296553



Patient 296528

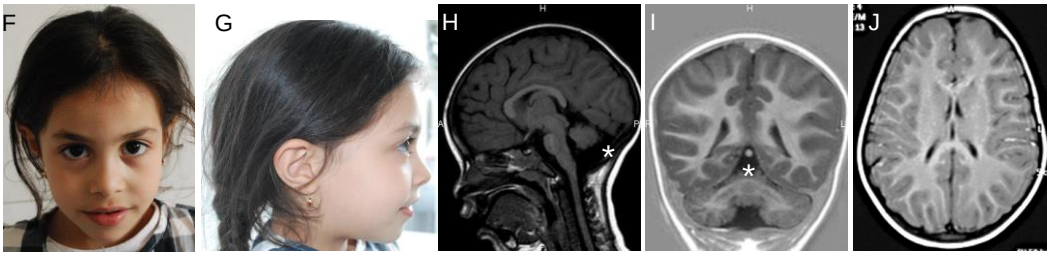
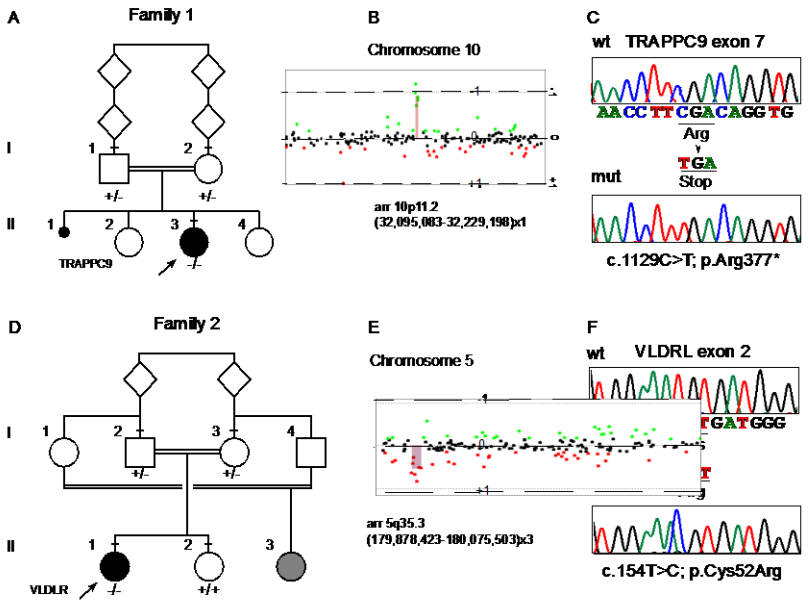


Figure 2



544

545

546

547

548

549

550

551

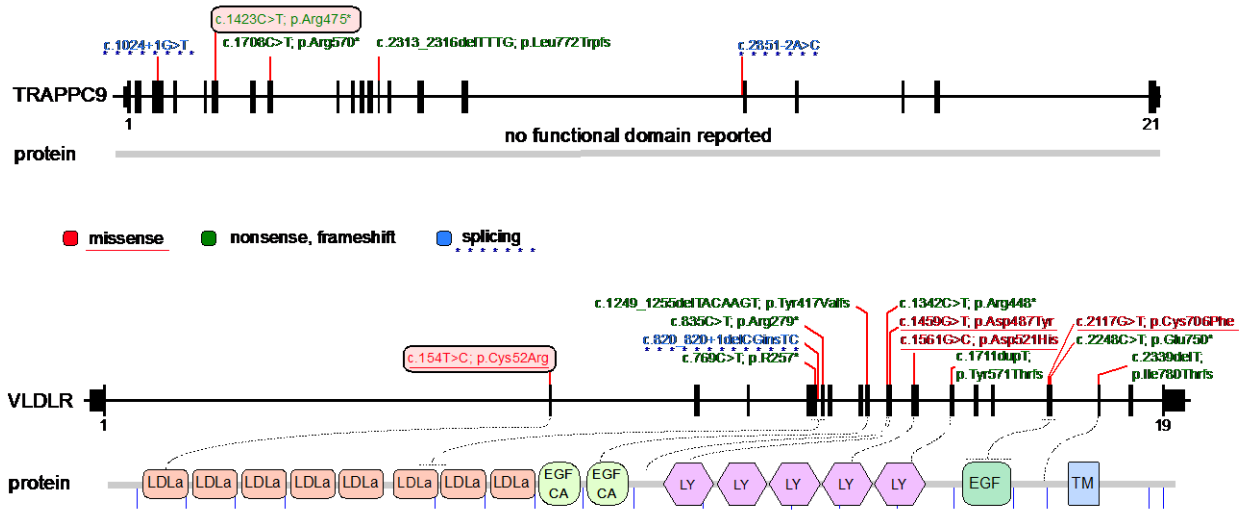
552

553

554

555

Figure 3



556

557

558

559

560

561

562

563

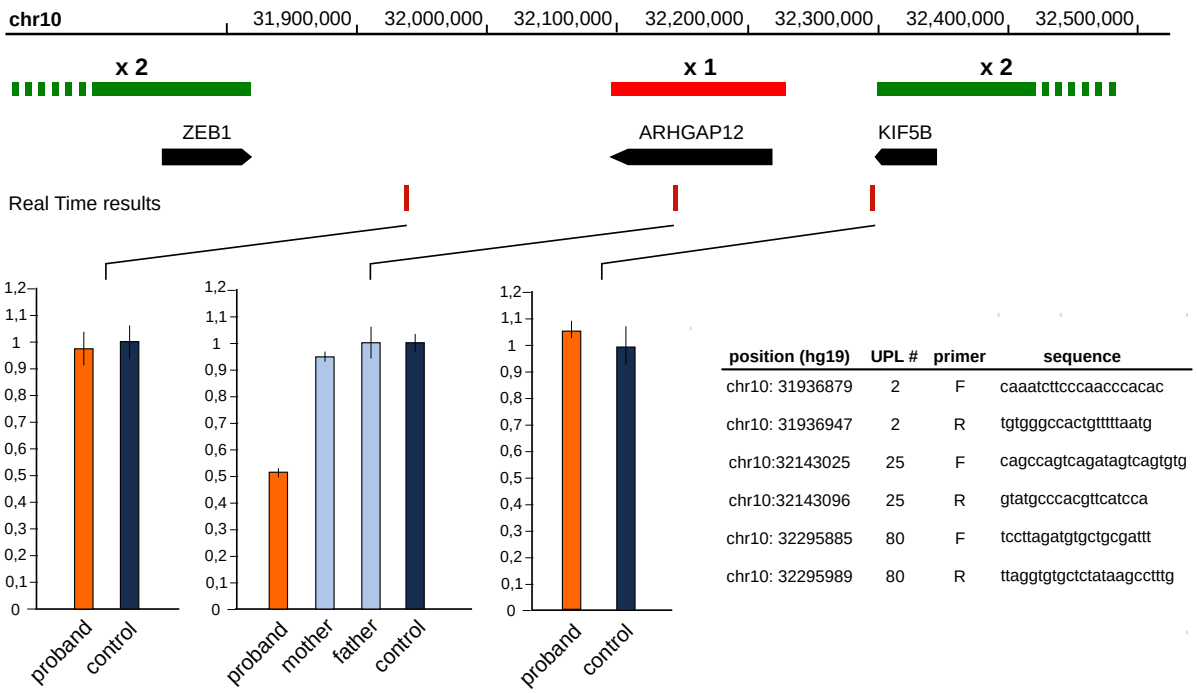
564

565

566

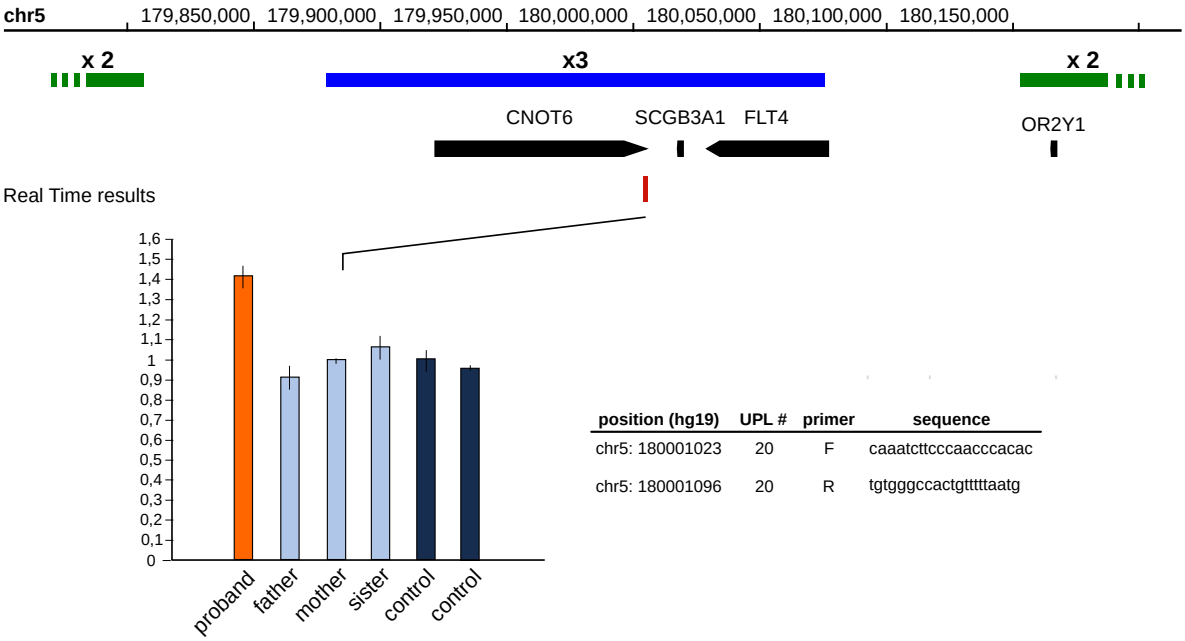
567

CASE 296553



position (hg19)	UPL #	primer	sequence
chr10: 31936879	2	F	caaatcttcccaacccacac
chr10: 31936947	2	R	tgtgggccactgtttttaatg
chr10:32143025	25	F	cagccagtcagatagtcagtgtg
chr10:32143096	25	R	gtatgcccacgttcatcca
chr10: 32295885	80	F	tccttagatgtgctgcgattt
chr10: 32295989	80	R	ttaggtgtgctctataagcctttg

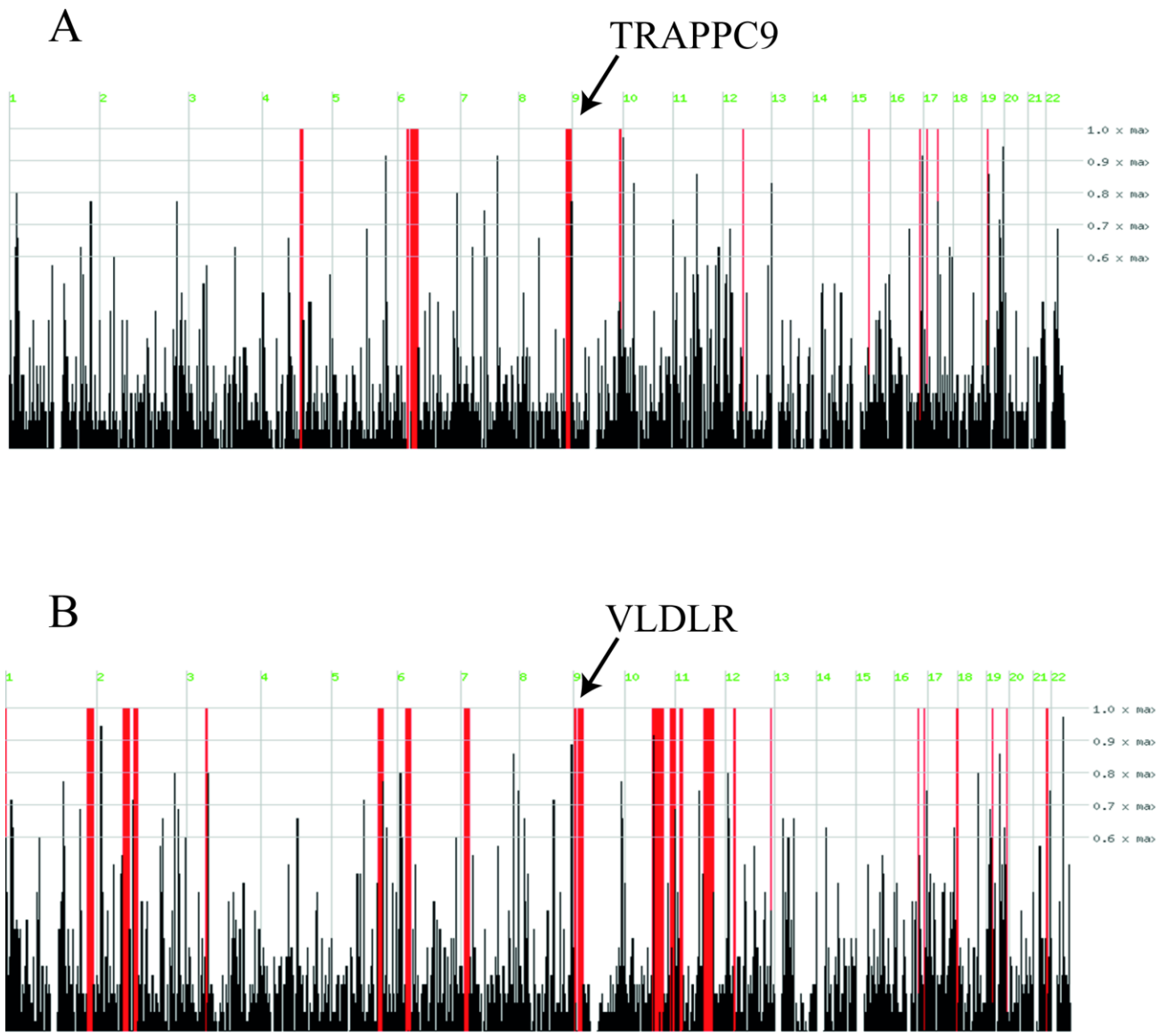
CASE 296528



position (hg19)	UPL #	primer	sequence
chr5: 180001023	20	F	caaatcttcccaacccacac
chr5: 180001096	20	R	tgtgggccactgtttttaatg

Suppl Fig.1: Real-time PCR analysis to confirm a-CGH results. For each patients, array-CGH are reported with the red bar indicating the minimal deleted region and the blue bar the minimal duplicated region. Flanking green bars represent regions with normal a-CGH signals. Genes spanning the rearrangement are shown as black (within the rearrangement) and grey (outside the rearrangement) arrows. The position of real-time PCR assays (UPL probe assay, Roche Diagnostics, Mannheim, Germany) used to validate a-CGH results are represented by vertical red bars. Histograms show the result for each assay (see flanking table for conditions). In both cases, real-time PCR showed the rearrangement was *de novo*. In case 296553, the deletion involved the entire *ARHGAP12* gene. In case 296528 the uncertainty in the duplication definition did not allow to estabilish if the upstream region of the *FLT4* gene was included.

568


A
TRAPPC9
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
1.0 x max
0.9 x max
0.8 x max
0.7 x max
0.6 x max
B
VLDLR
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
1.0 x max
0.9 x max
0.8 x max
0.7 x max
0.6 x max

```
                    40        50        60              70
H.sapiens         CEP-SQFQCTNGRCITLLWKCDGDEDCVDGSDEK----NC--VKKTCAESDFVCNNG
P.troglodytes     CEP-SQFQCTNGRCITLLWKCDGDEDCVDGSDEK----NC--VKKTCAESDFVCNNG
M.mulatta         CEP-SQFQCTNGRCITLLWKCDGDEDCVDGSDEK----NC--VKKTCAESDFVCNNG
C.lupus           CEP-SQFQCTNGRCITLLWKCDGDEDCADGSDEK----NC--VKKTCAESDFVCNNG
M.musculus        CDS-SQFQCTNGRCITLLWKCDGDEDCADGSDEK----NC--VKKTCAESDFVCKNG
R.norvegicus      CDS-SQFQCTNGRCITLLWKCDGDEDCTDGSDEK----NC--VKKTCAESDFVCKNG
B.taurus          CEA-NQFQCTNGRCITLLWKCDGDEDCTDGSDEK----NC--VKKTCAESDFVCNNG
X.tropicalis      CEG-SQFQCANGHCITSLWKCDGDEDCSDGSDES----SC--VKKTCAETDFVCNNG
G.gallus          CEE-SQFQCSNGRCIPLLWKCDGDEDCSDGSDES----AC--VKKTCAESDFVCNSG
D.rerio           CEQ-SQFQCGNGRCIPSVWQCDGDMDCSDGSDET----SC--VRKTCAEVDFVCRSG
D.melanogaster    CDE-KQFQCSTGECIPIRFVCDGSSDCPDHSDER--LEECKFTESTCSQEQFRCGNG
C.elegans         CDATNSFQCQDGRCIPMSWRCDGDIDCQNEEDEK----NCP-ISEVCGAEEHKCGEV
```

Supplementary material

Whole-exome sequencing

WES statistics are reported in supplementary table 1. Data annotation predicted 12,859 (case 296553) and 12,476 (case 296528) high-quality variants having functional impact (i.e., non-synonymous and splice site changes). Among them, 2,353 and 2,134 private, rare (minor allele frequency < 0.001) or clinically associated changes were further analyzed. Variants were filtered to retain rare or private homozygous sequence changes located within LoH regions, and *in silico* analyses of the predicted functional impact of individual variants and biological relevance of the encoded proteins allowed to identify an excellent disease gene candidate in each patient (Supplementary table 2 and supplementary figure 3).

Supplementary table 1. WES data output.

Cases	**296553**	**296528**
Total number of reads	70,139,440	125,619,358
Mean read length (bp)	90	90
Target regions coverage (%)[1]	98.8	99.1
Target regions coverage >10x (%)[1]	95.4	95.8
Average depth on target	56x	63x
Total variants	60,809	63,336
Variants with predicted effect on CDS[2]	12,476	12,859
- Novel variants, annotated variants (dbSNP141) with clinical association, minor allele frequency < 0.001, or unknown frequency[3]	453	574
- Homozygous variants	23	12
- Homozygous variants in LoH regions	15	6

[1] Nimblegen SeqCap EZ Library v 3.0

[2] Non synonymous SNPs and indels within the coding sequence and splice sites (±5 bases)

[3] All variants having a frequency < 0.05 in our in-house database.

Supplementary table 2.
List of the identified non synonymous homozygous variants located within LoH regions.

Patient/ gene	Position	Ref allele	Var allele	Predicted change	Novel/ annotated	Meta SVM score[1]	CADD score[1]	GeneDistiller overall score[2]
296553								
TRAPPC9	chr8:141407724	G	A	R377*	rs267607136	n.a.	44	121.8
ALDH5A1	chr6:24533797	A	G	M489V	.	0.4773	26.2	111.7
HFE	chr6:26093125	G	A	E171K	rs140080192	-1.1483	27.3	62.7
ICK	chr6:52874338	G	A	S507L	.	-1.1001	23.8	82.9
C6orf223	chr6:43970503	C	CGCG	A124AA	.	n.a.	14.14	0.0
HIST1H1A	chr6:26018004	G	A	c.-44C>T	rs201609154	n.a.	3.053	47.7
296528								
VLDLR	chr9:2635524	T	C	C52R	.	0.9023	33	2731.7
NAV2	chr11:20125247	C	A	N1271K	.	-0.6862	33	91.6
KIAA0020	chr9:2812275	G	T	H453N	.	-0.583	29.9	126.6
COG2	chr1:230810785	A	G	N314S	.	-1.141	26.3	136.9
BTN2A2	chr6:26392984	C	T	A244V	rs147634987	-0.8749	26	88.6
PTPLAD2	chr9:21026598	C	A	L89F	.	-1.0604	24.5	3.0
WDR11	chr10:122610998	C	G	H22Q	rs138044064	-0.5655	23.3	128.3
RNF103	chr2:86831304	C	T	E570K	.	-1.034	14.33	115.3
M1AP	chr2:74842194	T	C	Q108R	rs143027724	-0.9984	13.17	14.0
KDM4D	chr11:94731756	G	A	R407H	rs201511454	-1.0378	11.84	-11.2
ANKS1A	chr6:34952896	T	A	D350E	rs141760971	-1.0174	11.33	36.4
TRIM38	chr6:25972331	A	G	c.738+4 A>G	rs199757564	n.a	8.712	86.1
ZNF784	chr19:56133220	C	G	G290A	rs369499131	-0.9668	8.218	11.0
C2orf78	chr2:74043634	T	A	S762T	.	-1.0487	6.551	0.0
HLA-A	chr6:29911114	G	T	R138L	.	-0.9524	3.79	131

[1]Variants with scores <0 (dbNSFP) or <15 (CADD), predicting a negligible impact of the sequence change on protein structure and function, are highlighted in grey.

[2]GeneDistiller scoring (Seelow and others, 2008) used "focus on possible pathways" as prioritization method, and the following keywords for comparison with known genes: developmental delay, intellectual disability, mental retardation, microcephaly and motor stereotypies (case 296553); intellectual disability, mental retardation, ataxia and hypotonia (case 296528).