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

This is a pre print version of the following article:

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/1589161 since 2017-12-02T23:04:00Z

Published version:
DOI:10.1002/ajmg.a.37649

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
This is the author's final version of the contribution published as:

Giorgio, Elisa; Ciolfi, Andrea; Biamino, Elisa; Caputo, Viviana; Di Gregorio, Eleonora; Belligni, Elga Fabia; Calcia, Alessandro; Gaidolfi, Elena; Bruselles, Alessandro; Mancini, Cecilia; Cavalieri, Simona; Molinatto, Cristina; Cirillo Silengo, Margherita; Ferrero, Giovanni Battista; Tartaglia, Marco; Brusco, Alfredo. Whole exome sequencing is necessary to clarify ID/DD cases with de novo copy number variants of uncertain significance: Two proof-of-concept examples. AMERICAN JOURNAL OF MEDICAL GENETICS. PART A. 170 (7) pp: 1772-1779. DOI: 10.1002/ajmg.a.37649

The publisher's version is available at:
http://doi.wiley.com/10.1002/ajmg.a.37649

When citing, please refer to the published version.

Link to this full text:
http://hdl.handle.net/
Exome sequencing uncovers biallelic mutations in \textit{TRAPPC9} and \textit{VLDLR} and solve two syndromic intellectual disability cases with de novo CNVs.

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

\textsuperscript{1} University of Torino, Department of Medical Sciences, Turin, 10126, Italy
\textsuperscript{2} Department of Hematology, Oncology and Molecular Medicine, Istituto Superiore di Sanità, Rome, 00161, Italy
\textsuperscript{3} University of Torino, Department of Public Health and Pediatrics, Turin, 10126, Italy
\textsuperscript{4} Department of Experimental Medicine, Sapienza University of Rome, Rome, 00161, Italy
\textsuperscript{5} Città della Salute e della Scienza University Hospital, Medical Genetics Unit, Turin, 10126, Italy
\textsuperscript{6} Centro Diagnostico Cernaia, Magnetic Resonance Unit, Turin, 10122, Italy
\textsuperscript{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, \textit{de novo} CNV, intellectual disability, \textit{VLDRL}, \textit{TRAPPC9}. 
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.homozygosimapper.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(31,821,746x2,32,095,083-32,229,198x1,32,300,151), 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 ARGAP12, 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. Cys52 is predicted to be involved in an intramolecular disulfide bond with Cys67 (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 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


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.


Figure 1

Patient 296553

Patient 296528
Figure 2

A  Family 1

B  Chromosome 10

C  TRAPPC3 exon 7

D  Family 2

E  Chromosome 5

F  VLDRL exon 2

---

22
CASE 296553

chr10  31,900,000  32,000,000  32,100,000  32,200,000  32,300,000  32,400,000  32,500,000

Real Time results

Suppl Fig.1: Real-time PCR analysis to confirm a-CGH results. For each patient, 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 establish if the upstream region of the FLT4 gene was included.
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.

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

<sup>1</sup> Nimblegen SeqCap EZ Library v 3.0

<sup>2</sup> Non synonymous SNPs and indels within the coding sequence and splice sites (±5 bases)

<sup>3</sup> 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.

<table>
<thead>
<tr>
<th>Patient/gene</th>
<th>Position</th>
<th>Ref allele</th>
<th>Var allele</th>
<th>Predicted change</th>
<th>Novel/annotated</th>
<th>Meta SVM score(^1)</th>
<th>CADD score(^1)</th>
<th>GeneDistiller overall score(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>296553</td>
<td>chr8:141407724</td>
<td>G A</td>
<td>R377*</td>
<td>rs267607136</td>
<td>n.a.</td>
<td>44</td>
<td>121.8</td>
<td></td>
</tr>
<tr>
<td>TRAPPC9</td>
<td>chr6:24533797</td>
<td>A G</td>
<td>M489V</td>
<td>.</td>
<td>0.4773</td>
<td>26.2</td>
<td>111.7</td>
<td></td>
</tr>
<tr>
<td>ALDH5A1</td>
<td>chr6:26093125</td>
<td>G A</td>
<td>E171K</td>
<td>rs140080192</td>
<td>1.1483</td>
<td>27.3</td>
<td>62.7</td>
<td></td>
</tr>
<tr>
<td>HFE</td>
<td>chr5:52874338</td>
<td>G A</td>
<td>S507L</td>
<td>.</td>
<td>1.1001</td>
<td>23.8</td>
<td>82.9</td>
<td></td>
</tr>
<tr>
<td>ICK</td>
<td>chr6:43970503</td>
<td>C GCG G</td>
<td>A124AA</td>
<td>.</td>
<td>n.a.</td>
<td>14.14</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>C6orf223</td>
<td>chr6:26018004</td>
<td>G A</td>
<td>c.-44C&gt;T</td>
<td>rs201609154</td>
<td>n.a.</td>
<td>3.053</td>
<td>47.7</td>
<td></td>
</tr>
<tr>
<td>296528</td>
<td>chr9:2635524</td>
<td>T C</td>
<td>C52R</td>
<td>.</td>
<td>0.9023</td>
<td>33</td>
<td>2731.7</td>
<td></td>
</tr>
<tr>
<td>VLDLR</td>
<td>chr11:20125247</td>
<td>C A</td>
<td>N1271K</td>
<td>.</td>
<td>0.6862</td>
<td>33</td>
<td>91.6</td>
<td></td>
</tr>
<tr>
<td>NAV2</td>
<td>chr9:2812275</td>
<td>G T</td>
<td>H453N</td>
<td>.</td>
<td>-0.583</td>
<td>29.9</td>
<td>126.6</td>
<td></td>
</tr>
<tr>
<td>KIAA0020</td>
<td>chr1:230810785</td>
<td>A G</td>
<td>N314S</td>
<td>.</td>
<td>-1.141</td>
<td>26.3</td>
<td>136.9</td>
<td></td>
</tr>
<tr>
<td>COG2</td>
<td>chr6:26392984</td>
<td>C T</td>
<td>A244V</td>
<td>rs147634987</td>
<td>-0.8749</td>
<td>26</td>
<td>88.6</td>
<td></td>
</tr>
<tr>
<td>BTN2A2</td>
<td>chr9:21026598</td>
<td>C A</td>
<td>L89F</td>
<td>.</td>
<td>1.0604</td>
<td>24.5</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>PTPAD2</td>
<td>chr10:122610998</td>
<td>C G</td>
<td>H22Q</td>
<td>rs138044064</td>
<td>-0.5655</td>
<td>23.3</td>
<td>128.3</td>
<td></td>
</tr>
<tr>
<td>WDR11</td>
<td>chr2:86831304</td>
<td>C T</td>
<td>E570K</td>
<td>.</td>
<td>-1.034</td>
<td>14.33</td>
<td>115.3</td>
<td></td>
</tr>
<tr>
<td>RNF103</td>
<td>chr2:74842194</td>
<td>T C</td>
<td>Q108R</td>
<td>rs143027724</td>
<td>-0.9984</td>
<td>13.17</td>
<td>14.0</td>
<td></td>
</tr>
<tr>
<td>M1AP</td>
<td>chr7:94731756</td>
<td>G A</td>
<td>R407H</td>
<td>rs201511454</td>
<td>-1.0378</td>
<td>11.84</td>
<td>-11.2</td>
<td></td>
</tr>
<tr>
<td>KDM4D</td>
<td>chr6:34952896</td>
<td>T A</td>
<td>D350E</td>
<td>rs141760971</td>
<td>-1.0174</td>
<td>11.33</td>
<td>36.4</td>
<td></td>
</tr>
<tr>
<td>ANKS1A</td>
<td>chr3:56133220</td>
<td>C G</td>
<td>G290A</td>
<td>rs369499131</td>
<td>-0.9668</td>
<td>8.218</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>TRIM38</td>
<td>chr2:74043634</td>
<td>T A</td>
<td>S762T</td>
<td>.</td>
<td>-1.0487</td>
<td>6.551</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>ZNF784</td>
<td>chr6:29911114</td>
<td>G T</td>
<td>R138L</td>
<td>.</td>
<td>0.9524</td>
<td>3.79</td>
<td>131</td>
<td></td>
</tr>
<tr>
<td>HLA-A</td>
<td>chr3:56133220</td>
<td>C G</td>
<td>G290A</td>
<td>rs369499131</td>
<td>-0.9668</td>
<td>8.218</td>
<td>11.0</td>
<td></td>
</tr>
</tbody>
</table>

\(^1\)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).