Received: 31 March 2017 | Revised: 11 May 2017 | Accepted: 12 May 2017

DOI: 10.1111/cge.13060

WILEY CLINICAL GENETICS

ORIGINAL ARTICLE

Diagnostic application of a capture based NGS test for the concurrent detection of variants in sequence and copy number as well as LOH

A. Vetro[1] | D. Goidin[2] | I. Lesende[2] | I. Limongelli[3] | G.N. Ranzani[4] | F. Novara[1] | M.C. Bonaglia[5] | B. Rinaldi[1] | F. Franchi[6] | E. Manolakos[7,8] | F. Lonardo[9] | F. Scarano[9] | G. Scarano[9] | L. Costantino[10] | S. Tedeschi[10] | S. Giglio[11] | O. Zuffardi[1]

[1]Department of Molecular Medicine, University of Pavia, Pavia, Italy

[2]Diagnostics and Genomics Group, Agilent Technologies Inc., Santa Clara, California

[3]enGenome s.r.l, Pavia, Italy

[4]Department of Biology and Biotechnology, University of Pavia, Pavia, Italy

[5]Cytogenetics Laboratory, Scientific Institute IRCCS E. Medea, Lecco, Italy

[6]Laboratorio Genetica, Azienda Ospedaliera Arcispedale Santa Maria Nuova, Reggio Emilia, Italy

[7]Clinical Laboratory Genetics, Access to Genome, Athens, Greece

[8]Clinical Laboratory Genetics, Access to Genome, Thessaloniki, Greece

[9]U.O.S.D. Genetica Medica—A.O.R.N, Benevento, Italy

[10]Medical Genetics Laboratory, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy

[11]Medical Genetics Unit, Meyer Children's University Hospital, Firenze, Italy

Correspondence
Annalisa Vetro, Department of Molecular Medicine, University of Pavia, via Forlanini, 14—27100 Pavia, Italy.
Email: annalisa.vetro@unipv.it
Orsetta Zuffardi, Department of Molecular Medicine, University of Pavia, via Forlanini, 14—27100 Pavia, Italy.
Email: orsetta.zuffardi@unipv.it

Funding information
Telethon Italy Grant, Grant/Award number: GGP13060.

Whole exome sequencing (WES) has made the identification of causative SNVs/InDels associated with rare Mendelian conditions increasingly accessible. Incorporation of softwares allowing CNVs detection into the WES bioinformatics pipelines may increase the diagnostic yield. However, no standard protocols for this analysis are so far available and CNVs in non-coding regions are totally missed by WES, in spite of their possible role in the regulation of the flanking genes expression. So, in a number of cases the diagnostic workflow contemplates an initial investigation by genomic arrays followed, in the negative cases, by WES. The opposite workflow may also be applied, according to the familial segregation of the disease.

We show preliminary results for a diagnostic application of a single next generation sequencing panel permitting the concurrent detection of LOH and variations in sequences and copy number. This approach allowed us to highlight compound heterozygosity for a CNV and a sequence variant in a number of cases, the duplication of a non-coding region responsible for sex reversal, and a whole-chromosome isodisomy causing reduction to homozygosity for a WFS1 variant. Moreover, the panel enabled us to detect deletions, duplications, and amplifications with sensitivity comparable to that of the most widely used array-CGH platforms.

KEYWORDS

copy number variations, exome sequencing, genetic diagnosis, isodisomy, loss of heterozygosity

1 | BACKGROUND

The identification of the causative DNA lesion is central to genomic medicine, although in most cases it is complicated by genetic heterogeneity, variable expressivity, and incomplete penetrance. The emerging literature on digenic inheritance[1,2] makes further challenging the identification of disease-causing genes and variants.

The expectations created by precision medicine and the immense funding dedicated to it worldwide have made urgent to get the molecular diagnosis in genetic diseases, in order to entertain the more appropriate therapeutic strategy for the patient, and any possible prevention in relatives and fetuses at risk.

Since the last 10 years, genomic arrays have represented the first-tier analysis in different clinical conditions, both sporadic and familial,[3–5] allowing to detect Copy Number Variations (CNVs) or, when SNP-specific probes are included in the platform, also regions of copy-neutral loss of heterozygosity (cnLOH). The latter may be responsible for a clinical condition if the affected region contains either imprinted genes or at least one disease-variant, heterozygous in 1 parent and reduced to homozygosity.[6,7]

More recently, Next Generation Sequencing (NGS) technologies have made increasingly accessible the identification of both single nucleotide variants (SNVs) and small insertions/deletions (InDels), largely overcoming the problem of genetic and phenotypic heterogeneity.[8]

However, among the 141 000 genomic lesions reported in the Human Gene Mutation Database (HGMD), 10% consists of the total or partial deletion or duplication of the disease-associated genes.[9] Moreover, the association of a CNV with a loss of function or hypomorphic SNV in the other allele accounts for a proportion of autosomal-recessive diseases.[10,11] Although whole-genome sequencing (WGS) is able to provide information regarding point mutations and structural abnormalities, the cost and the objective difficulty in interpreting the numerous variants impair its application on a routine basis.

In contrast, the easier handling of the whole exome sequencing (WES) data made it part of the diagnostic routine in many genetics laboratories. Incorporation of the CNVs detection in the WES bioinformatics pipelines is more and more common, although no standard protocols are so far available, and the accuracy of the CNVs call is influenced by several factors, including the panel design, the sequencing technology, the reads length, and the local sequence context.[12–14] Obviously, CNVs in non-coding regions are totally missed by WES, in spite of their possible role in the regulation of flanking genes expression.[15]

Therefore, in a number of laboratories an initial investigation by genomic arrays is followed, in the negative cases, by WES analysis or sequencing of specific panels of genes fitting with the proband phenotype.

Thus, achieving a molecular diagnosis often involves multiple different investigations that are planned case-by-case based on the more probable hypothesis. The availability of a single test, able to identify at the same time most of the molecular causes of the genetic disorders, would be of great advantage, allowing to reduce the number of tests, the final cost, and the reporting time-frame.

We conducted a validation study on a cohort of cases with previously identified genomic alterations, which was re-analyzed by using a commercial target-enrichment kit, allowing the concurrent detection of CNVs, SNVs/InDels and LOH events. In most of these cases, several tests had been applied before to reach a diagnosis, whereas by using this approach we would have been able to highlight the molecular etiology of the disease by a single experiment.

2 | MATERIALS AND METHODS

2.1 | Clinical details and previous molecular investigations

The details for the reported cases are summarized in Table 1. Cases 1, 11 and 12 were prenatally diagnosed because of ultrasound abnormalities. Cases 3, 4-6, 11, 13, 15, and 17 have been previously described.[16–22] Relevant data are available at DECIPHER (https://decipher.sanger.ac.uk/index, see Table 1).

2.1.1 | Case 1

A 41-year-old woman was referred to genetic investigations at 20 weeks of gestation because of ultrasound abnormalities including hypoplastic cerebellar vermis and corpus callosum agenesis. Conventional cytogenetics showed a rearranged chromosome 8p, further characterized by array-CGH (105 K) as: arr[hg19] 8p23.3p23.1 (73 810-6 914 026)x1, 8p23.1p11.1(12 557 768-43 527 906)x3 dn (Figure 1). A 4.2 Mb deletion of 18q, arr[hg19] 18q21.33q22.1 (60 572 379-64 745 957)x1 mat, was also highlighted (not shown).

2.1.2 | Case 2

The patient, a 2-year-old male, was suspected to be affected by the William-Beuren syndrome (MIM: #194050) because of supravalvular aortic stenosis, mild psychomotor delay, periorbital fullness and friendly behavior. Array-CGH (180 K) showed a 7q11.23 de novo deletion: arr[hg19] 7q11.23(72 726 578-74 139 331)x1 dn (not shown).

2.1.3 | Case 7

The newborn male was born spontaneously at 36 + 4 weeks of gestation (Apgar 8/1′ and 8/5′). Birth weight and length were 3330 g (25-50° centile) and 49.5 cm (25° centile), respectively and head circumference was 34.5 cm (50° centile). Bilateral arthrogryposis, involving both knees and elbows, and right cryptorchidism were noticed. Kidney and brain echography were normal, whereas echocardiography showed patent foramen ovale.

Two angiomas were noticed, a frontal one extended up to the eyelids, and a nuchal one. Karyotype analysis on peripheral blood lymphocytes, showed the presence in all the metaphases of a SMC (supernumerary marker chromosome) that was classified, after array-CGH (180 K, Figure 1) and FISH analyses, as a neocentric inv dup(17)(p13.3): 47,XY,+mar.arr[hg19] 17p13.3(51 885-1 879 066)x4 dn. As the average $\log_2$ratio the amplified region was 0.83, we could not rule out that the marker was present in mosaic in the patient blood, at least in 80% of the cells.

TABLE 1 Details and molecular defects of the reported cases

id	Clinical condition	Known molecular defect and size of the CNV[a]	OneSeq analysis results and size of the CNV	Identified by Oneseq	Ref
Case 1	Fetus with hypoplastic cerebellar vermis and corpus callosum agenesis	arr[hg19] 8p23.3p23.1(73 810-6 914 026)x1,8p23.1p11.1 (12 557 768-43 527 906)x3 dn,18q21.33q22.1 (60 572 379-64 745 957)x1 mat; size: >30; 4.1 Mb	8p23.3p23.1(12 630-7 220 339)x1, 8p23.1p11.1 (12 171 110-43 756 290)x3, 18q21.33q22.1(60 593 880-64 776 663)x1; size: >30; 4.1 Mb	Y	
Case 2	Williams-Beuren syndrome [MIM #194050]	arr[hg19] 7q11.23(72 726 578-74 139 331)x1 dn; size: 1.4 Mb	7q11.23(72 707 404-74 138 389)x1; size: 1.4 Mb	Y	
Case 3	46,XX DSD, SRY negative, with ambiguous genitalia and ovotestis	arr[hg19] 17q24.3(69 404 081-69 872 909)x3 pat; size: 470 kb	17q24.3(69 433 066-69 877 499)x3; size: 440 kb	Y	Case 2 in Ref. 16; DECIPHER id: 293615
Case 4	Familial adenomatous polyposis [MIM #175100]	arr[hg19] 5q22.2(112 038 759-112 296 071)x1, 19q13.42 (53 966 153-54 405 781)x3; size: 260 and 440 kb	5q22.2(112 041 577-112 252 150)x1, 19q13.42 (53 932 628-54 406 554)x3; size: 211 and 470 kb	Y	Case 6 in Ref. 17; DECIPHER id: 337180
Case 5	Familial adenomatous polyposis [MIM #175100]	arr[hg19] 5q22.2(112 174 106-112 197 427)x1; size: 23 kb	5q22.2(1 271 792-1 202 336) x1; size: 30 kb	Y[b]	Case 4 in Ref. 17; DECIPHER id: 340989
Case 6	Carney complex [MIM #160980] and SMC constituted by 2 extra copies of 1p31.1 (amplification)	arr[hg19] 1p31.1(83 288 652-84 890 437)x4 dn; size: 1.6 Mb	1p31.1(83 279 478-84 917 765)x4; size: 1.6 Mb	Y	Ref. 18; DECIPHER id: 340991
Case 7	Facial dysmorphism, arthrogryposis, cryptorchidism, karyotype: 47,XY,+mar	arr[hg19] 17p13.3(51 885-1 879 066)x4 dn; size: 1.8 Mb	17p13.3(449-1 880 205)x4; size: 1.8 Mb	Y	DECIPHER id: 340992
Case 8	Intellectual disability, karyotype: 46,XX,del(2)/46, XX,der(2)t(2,14)	arr[hg19] 2q37.3(241 591 565-243 087 697) x1,14q24.3q32.33 (78 504 178-107 287 446) x2-3 dn; size: 1.5 and 28.7 Mb	2q37.3(241 558 491-243 016 443)x1, 14q24.3q32.33(78 566 512-106 950 884)x2 ~ 3; size: 1.5 and 28.3 Mb	Y	
Case 9	46,XX[15]/47,XX,+22[31]	arr[hg19] (22)x2∼3; size: >30 Mb	22q11.1q13.33(17 253 851-51 176 309) x2-3; size: >30 Mb	Y	
Case 10	46,X,idic(Y)(q11.22)[31]/45,X [6]	arr[hg19] Xp22.33 or Yp11.32 (61 091-2 689 364 or 11 091-245 925) x2 ≈ 3,Yp11.31q11.22 (2 650 450-19 511 623) x2 ≈ 3,Yq11.22(19 538 856-59 335 869)x0; size: 16.9 and >39 Mb	Yp11.31q11.22(2 650 695-19 566 048)x2-3, Yq11.22 (20 808 947-58 986 878)x0; size: 16.9 and >39 Mb	Y	
Case 11	Fetus with abnormal ultrasound and suspected TAR syndrome [MIM #274000]	arr[hg19] 1q21.1(145 413 388-145 747 269)x1, size: 330 kb, and NM_005105.4: c.-21G>A	1q21.1(145 382 958-145 829 536)x1, size: 450 kb, NM_005105.4 (RBM8A):c.-21G>A**	Y	Ref. 19
Case 12	Fetus with abnormal ultrasound and suspected TAR syndrome [MIM #274000]	arr[hg19] 1q21.q(145 382 387-145 833 025)x1, size: 450 kb, and NM_005105.4: c.-21G>A	1q21.1(145 378 288-145 829 536)x1, size: 450 kb, NM_005105.4 (RBM8A):c.-21G>A**	Y	DECIPHER id: 340424
Case 13	Bartter syndrome, type 3 [MIM #607364]	arr[hg19] 1p36.13(16 364 253-16 387 572)x1, size: 23 kb, and NM_000085.3: c.1101G>A (p.T367*)	1p36.13(16 370 886-16 383 605)x1, size: 13 kb, NM_000085.3(CLCNKB): c.1101G>A (p.T367*)	Y	5BE99 in Ref. 20
Case 14	Bartter syndrome, type 3 [MIM #607364]	NM_004070.3(*CLCNKA*):c.(576+1_577-1)_(2016+1_2017-1) del; NM_000085.3 (CLCNKB):c.(?_-136)_(576+1_577-1)del; c.446T>A(p.Val149Glu)	1p36.13(16 354 404-16 375 449)x1, size: 21 kb, NM_000085.3(CLCNKB): c.446T>A(p.Val149Glu)	Y	DECIPHER id: 341010
Case 15	Cystic Fibrosis [MIM #219700]	arr[hg19] 7q31.2(117 176 613-117 243 758)x3, size: 67 kb, and NM_000492.3(CFTR): c.1521_1523delCTT	7q31.2(117 176 471-117 246 978)x3, size: 70 kb, NM_000492.3(CFTR): c.1521_1523delCTT	Y	Ref. 21

(*Continued*)

TABLE 1 Continued

id	Clinical condition	Known molecular defect and size of the CNV[a]	OneSeq analysis results and size of the CNV	Identified by Oneseq	Ref
Case 16	Retinitis Pigmentosa [MIM: #268000]	arr[hg19] 1q31.33 (197 434 389-197 455 060) x0 pat mat, size: 21 kb	1q31.33(197 420 690-197 458 350)x0, size: 38 kb	Y[b]	DECIPHER id: 341013
Case 17	Wolfram syndrome [MIM #222300]	NM_001145853.1: c.1348_1350delinsTAG (p. His450*)	c.1348_1350delinsTAG (p. His450*) and chr4:49 792-190 862 155 LOH	Y	Ref. 22

Ref, Reference for the cases previously described

[a] For array-CGH results the defect is reported according to the ISCN nomenclature.

[b] The aberration was not detected by using standard parameters, but identified by a dedicated analysis.

2.1.4 | Case 8

This 5-year-old female patient was referred because of intellectual disability. Conventional cytogenetics on peripheral blood lymphocytes revealed a de novo unbalanced translocation with a derivative chromosome 2 having a portion of chromosome 14q attached to its distal long arm: der(2)t(2;14)(q27;q24.3). FISH analysis with a 2q subtelomeric probe (Vysis, Abbott, Abbott Park, Illinois) showed a mosaic condition with a simple 2q terminal deletion in 70 metaphases and a 2q deletion in a derivative chromosome der(2)t(2,14) in 50 metaphases. Array-CGH (180 K) showed the presence of a homogeneous 2q distal deletion whereas the 14q24.3-qter region was characterized by a $\log_2$ ratio of +0.44, suggestive of a mosaic duplication involving about 70% of the DNA from blood: arr[hg19] 2q37.3(241 591 565-243 087 697)x1,14q24.3q32.33(78 504 178-107 287 446)x2-3 (Figure S1, Supporting information).

2.1.5 | Case 9

The patient was ascertained at the age of 12 years because of severe scoliosis. Her growth parameters have always been below the third centile, and facial and body asymmetry with right hemibody hyperplasia were evident. She showed facial dysmorphisms, renal fusion with discoid kidney and ostium secundum atrial septal defect, surgically treated at 7 years of age. She also showed hyper- and hypopigmented skin lines.

Cytogenetic investigations were performed on both peripheral blood lymphocytes and cultured skin fibroblasts. The blood karyotype resulted to be 46,XX, whereas a trisomy 22 was found in two-third of the fibroblast metaphases: 46,XX[15]/47,XX + 22[31]. Array-CGH (180 K) from skin fibroblasts showed the trisomy 22 with an average $\log_2$ ratio of +0.48 (Figure S1).

2.1.6 | Case 10

A 7-year-old boy was referred to our unit to re-examine a rearrangement detected at prenatal cytogenetic investigations. He presented with normal auxological parameters but a mild language delay requiring the support of a speech therapist.

Fetal karyotype was 46,X,idic(Y)(q11.2).ish idic(Y)(q11)(DYZ3++,SRY++)[18]/45,X[4]/46,X,r(Y).ish r(Y). Conventional and molecular cytogenetics (array-CGH 180 K, Figure S1) on peripheral blood resulted: 46,X,idic(Y)(q11.2)[31]/45,X[6].arr[hg 19] Yp11.32q11.221 (11 091-19 511 263)x1 ~ 2,Yq11.222q12(19 538 856-59 335 869)x0.

2.1.7 | Case 12

A 24-year-old female, presented at 15 weeks of pregnancy because of fetal phocomelia of the upper limbs, IUGR, bilateral clubfoot, hypoplastic nasal bone. The couple opted for a voluntary termination of pregnancy.

Conventional cytogenetics on amniocytes showed a 46,XY karyotype. In the suspect of Roberts syndrome [MIM: #268300] the *ESCO2* gene was analyzed by Sanger sequencing showing normal results. Array-CGH analysis (Cytochip 4 × 180 K BluGnome, Cambridge, UK) highlighted a ≈ 450 Kb de novo deletion of chromosome 1q21.1 including *RBM8A*: arr[hg19] 1q21.q(145 382 386-145 833 024)x1 (Figure 2), thus suggesting a diagnosis of TAR syndrome [MIM: #274000].

2.1.8 | Case 13

The patient [5BE99 in Ref. 20] was a 6-month-old girl, presenting with growth retardation, poor appetite, polyuria, and vomit. Blood and urine analysis suggested a salt waste and kidney lithiasis, confirmed by echography. A tentative diagnosis of Bartter syndrome was performed and genetic investigations identified compound heterozygosity for a *CLCNKB* c.1101G>A missense mutation and 1p36 deletion of approximately 23 kb removing the entire gene. The latter was detected by MLPA and further characterized by array-CGH (244 K): arr[hg19] 1p36.13(16 364 253-16 387 572)x1 (Figure 3).

2.1.9 | Case 14

The patient was born at 37 weeks of gestation from a pregnancy characterized by polyhydramnios. Immediately after birth, he received endotracheal intubation, manual ventilation and oxygen therapy with recovery of the general conditions. He presented with short stature, axial hypotonia, vomiting and polyuria. Kidney echography was normal.

A tentative diagnosis of Bartter syndrome was made. Genetic analysis identified compound heterozygosity for a missense c.446T>A mutation in *CLCNKB* and a deletion encompassing *CLCNKA* exons 7 to 19 and *CLCNKB* exons 1 to 6 [NM_004070.3:c.(576+1_577−1)_(2016 + 1_2017−1)del; NM_000085.3:c.(?_-136)_(576+1_577−1)del], the latter detected by using MLPA.

2.1.10 | Case 16

This child, born from consanguineous parents, was referred because of retinitis pigmentosa (RP, MIM: #268000) at the age of 4 years.

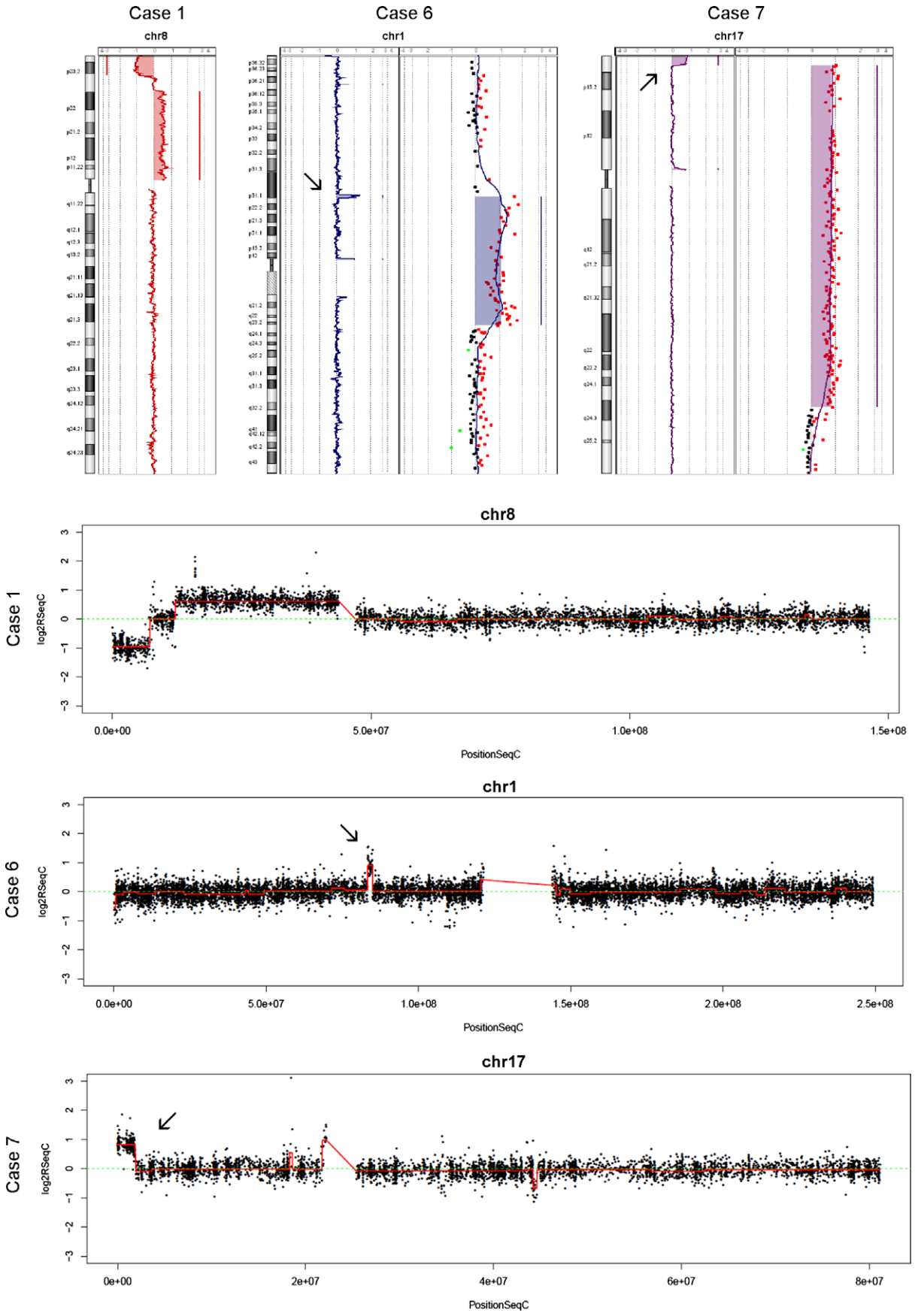

FIGURE 1 Array-CGH and sequencing results of cases with large rearrangements or supernumerary marker chromosomes. Upper panels, from left to right: array-CGH profiles of the inv dup del(8p) in case 1, the amplification of 1p31.1 in case 6,[18] and of 17p in case 7, both indicated by arrows. For cases 6 and 7, an enlargement of the amplified region is provided, highlighting the $\log_2$ ratio average of +1. The lower panels show the copy number profiles of the chromosomes of interest from the same cases, as analyzed by EXCAVATOR. Arrows indicate the amplifications at 1p31.1 (case 6) and 17p13.3 (case 7). The $\log_2$ratio scale is shown on the left for each panel

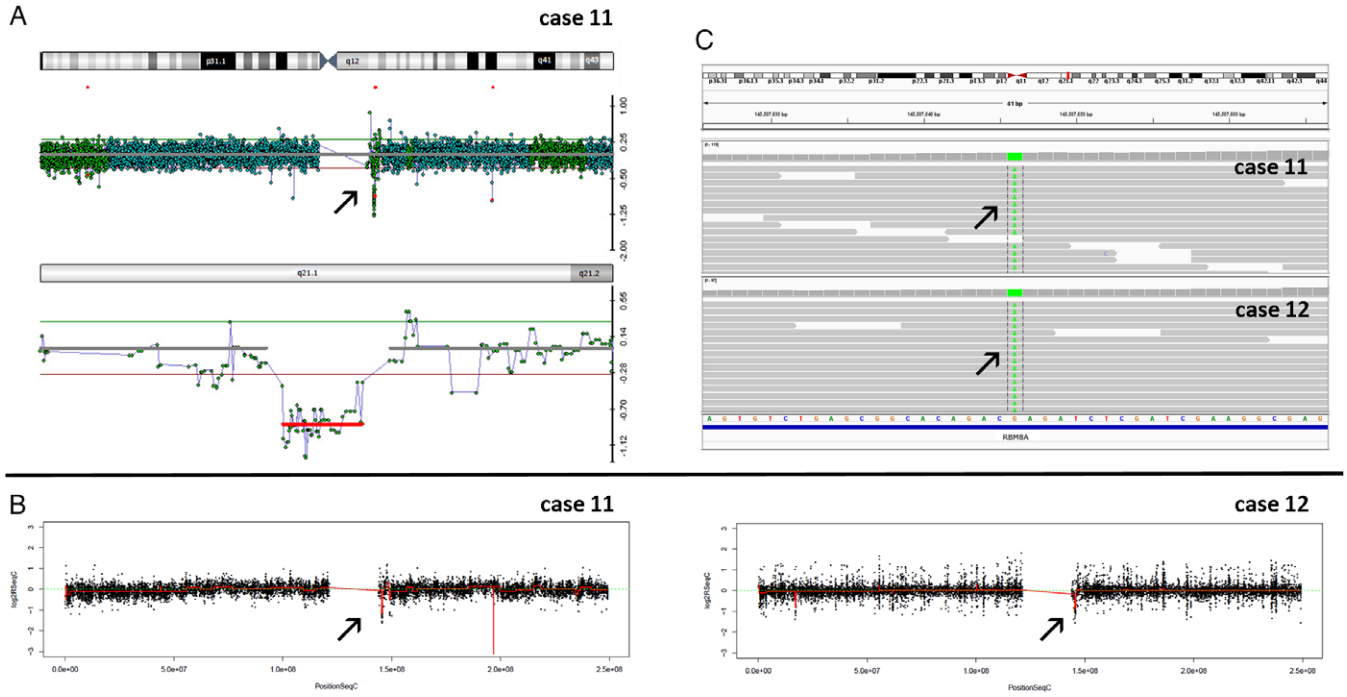

FIGURE 2 Array-CGH and sequencing results of the 2 fetuses with TAR syndrome. A, array-CGH profile (Cytochip, BluGnome) of chromosome 1 in case 12 shows an approximate 450 kb deletion of 1q21.1 (upper panel, arrow) involving a number of genes, including *RBM8A*. An enlargement of the deleted region is provided in the lower panel. A similar deletion was observed in case 11, whose array profile is illustrated in Papoulidis et al[19]; B, Copy number profile of chromosome 1 from both cases as analyzed by EXCAVATOR; C, IGV screenshot shows the c.-21G>A variant of *RMB8A* in both cases (in green, arrows). Grey bars represent the mapped reads aligned to the reference genome, whose sequence is shown in the bottom (colored letters). A coverage plot is displayed in the upper part of each panel

Sequencing of a panel of genes related to this condition resulted negative. Array-CGH with a 180 K platform showed a homozygous deletion of 1q including the RP gene *CRB1*: arr[hg19] 1q31.33 (197 434 389-197 455 060)x0* pat mat (Figure 4C).

2.2 | Conventional and molecular cytogenetic analyses

We set up PHA-stimulated lymphocytes cultures of cases 3, 6-10, and skin fibroblasts cultures of cases 6, 8 and 9. Karyotype analysis was performed on GTG-banded metaphases according to standard procedures. For cases 1, 11 and 12, conventional and molecular cytogenetic analysis was performed on fetal samples (cultured amniocytes for cases 1, 11 and chorion villi for case 12) according to standard procedures.

Array-CGH analysis was performed in all cases by using standard 105, 180 or 244 K arrays (Agilent Technologies, Santa Clara, California), as reported elsewhere.[23] For case 12 a different platform was used, according to manufacturer's instructions (Cytochip 4 × 180 K BluGnome).

2.3 | Sanger sequencing and MLPA analysis

Conventional capillary sequencing was applied in cases 11 to 15 and 17, by using standard methods to analyze the gene/s of interest, as reported elsewhere.[19–22] Primers and PCR conditions are available on request. MLPA analysis was applied in cases 4, 5 and 13 to 15 by using commercial kits (MRC-Holland, Amsterdam, The Netherlands), as described.[17,20] Case 15 was investigated by different techniques, as described.[21]

2.4 | OneSeq Constitutional Research Panel Design

OneSeq Constitutional Research Panel (Agilent Technologies) is a target-enrichment assay developed to detect simultaneously genome-wide CNVs and cnLOH as well as coding SNVs/InDels by NGS. The 28 Mb design includes: (1) the SureSelect Focused Exome baits (16 Mb design targeting disease-associated genes from HGMD, OMIM and ClinVar), (2) a set of genome-wide backbone baits with a median probe space of 50 Kb, giving a 300 Kb functional copy number resolution, (3) a set of baits targeting genomic regions including SNPs with a MAF (minor allele frequency) higher than 0.2, allowing the detection of cnLOH at a minimum of 5 Mb resolution, and (4) a set of baits enabling a >25-50 Kb resolution in disease-associated ClinGen Regions.[24] Design files are available at https://earray.chem.agilent.com/suredesign.

2.5 | Library preparation and sequencing

gDNA was extracted from lymphocytes for all cases except 1, 11 and 12, where fetal DNA was extracted from cultured amniocytes or chorion villi. Library preparation and target enrichment were performed by using the SureSelectXT OneSeq Constitutional Research Panel kit (5190-8702, Agilent Technologies). The SureSelectXT low input

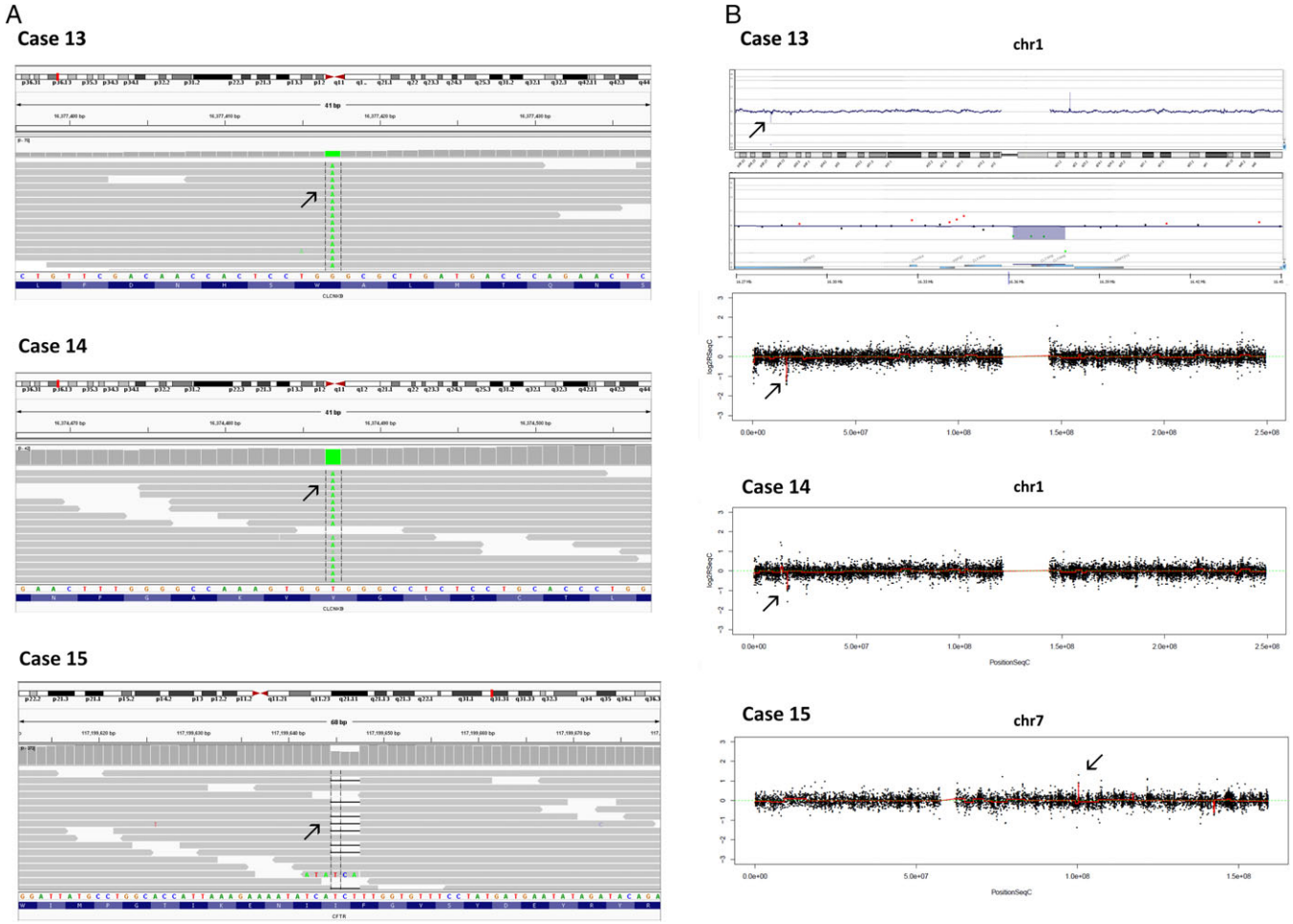

FIGURE 3 Results of the sequence analysis in compound heterozygous cases having a CNV and a SNV: Bartter syndrome (cases 13 and 14) and cystic fibrosis (case 15). A, IGV screenshots: the 2 heterozygous variants in *CLCNKB* detected in cases 13 and 14 (arrows, c.1101G>A and c.446T>A), respectively, and the heterozygous c.1521_1523delCTT variant of *CFTR* in case 15. B, The copy number profiles of chromosomes 1 and 7, with the CNVs highlighted by arrows: for case 13 the array-CGH and the EXCAVATOR profiles are shown, whereas for cases 14 and 15 the EXCAVATOR profile only is reported. In cases 13 and 14 Bartter syndrome is the consequence of a single nucleotide variant in *CLCNKB* and of a heterozygous deletion removing part of the *CLCNKA* and *CLNKB* genes on the other allele. In case 15, a large duplication, involving exons 6b to 16, disrupts the frame of the gene leading to the insertion of 8 amino-acid residues, followed by a premature stop codon, as reported elsewhere[21]

protocol version B.1 was adopted for all samples with few modifications: (1) DNA shearing was performed in 2 steps of 90 seconds (Duty Factor 10%, Peak Incident Power 175, Cycles per Burst 200) separated by a quick spin to collect droplets at the bottom of the Covaris microTube (Covaris, Woburn, Massachusetts); (2) half (15 μL) of the adaptor-ligated library was amplified for 8 cycles instead than 10 with the remaining half saved as a backup. Two reference DNAs, OneSeq Reference DNA Male, 5190-8848 and OneSeq Reference DNA Female, 5190-8850 (Agilent Technologies), were also prepared and used as controls.

After hybridization and indexing, 8-samples pools were prepared and sequenced by using a rapid run 2 × 100 bp paired-end protocol, on the HiSeq2500 (Illumina, San Diego, California).

2.6 | Bioinformatics analysis and variants interpretation

We used an in-house pipeline that implemented for CNVs analysis by EXCAVATOR.[25] Accordingly, reads were aligned to the human genome reference (GRCh37/hg19). Data processing and variant annotation was performed as previously reported.[26] SNVs and InDels were further filtered for those variants not overlapping the OneSeq Constitutional Research Panel targeted regions within padding regions of 200 bp at each target site. Selected SNVs/InDels were visually inspected by using the open-source tool Integrative Genomics Viewer IGV 1.2[27] to confirm the quality of the alignment.

For CNV detection, data were analyzed by EXCAVATOR, a publicly available tool, which uses $\log_2$ ratio of mapped read counts at the exon (or target region) level between case/control samples, even from different experiments, following a normalization for GC content, genomic mappability and exon size. An extended version of segmentation shifting level model (SLM), taking into account the distance between consecutive exons, combines nearby exons with a similar $\log_2$ ratio generating segments that are then classified following a 5-state classification scheme (2-copy deletion, 1-copy deletion, normal, 1-copy duplication and multiple-copy amplification).

The Agilent software SureCall (v. 3.0.1.4) was also used to analyze paired-end FASTQ files from all the samples, under default

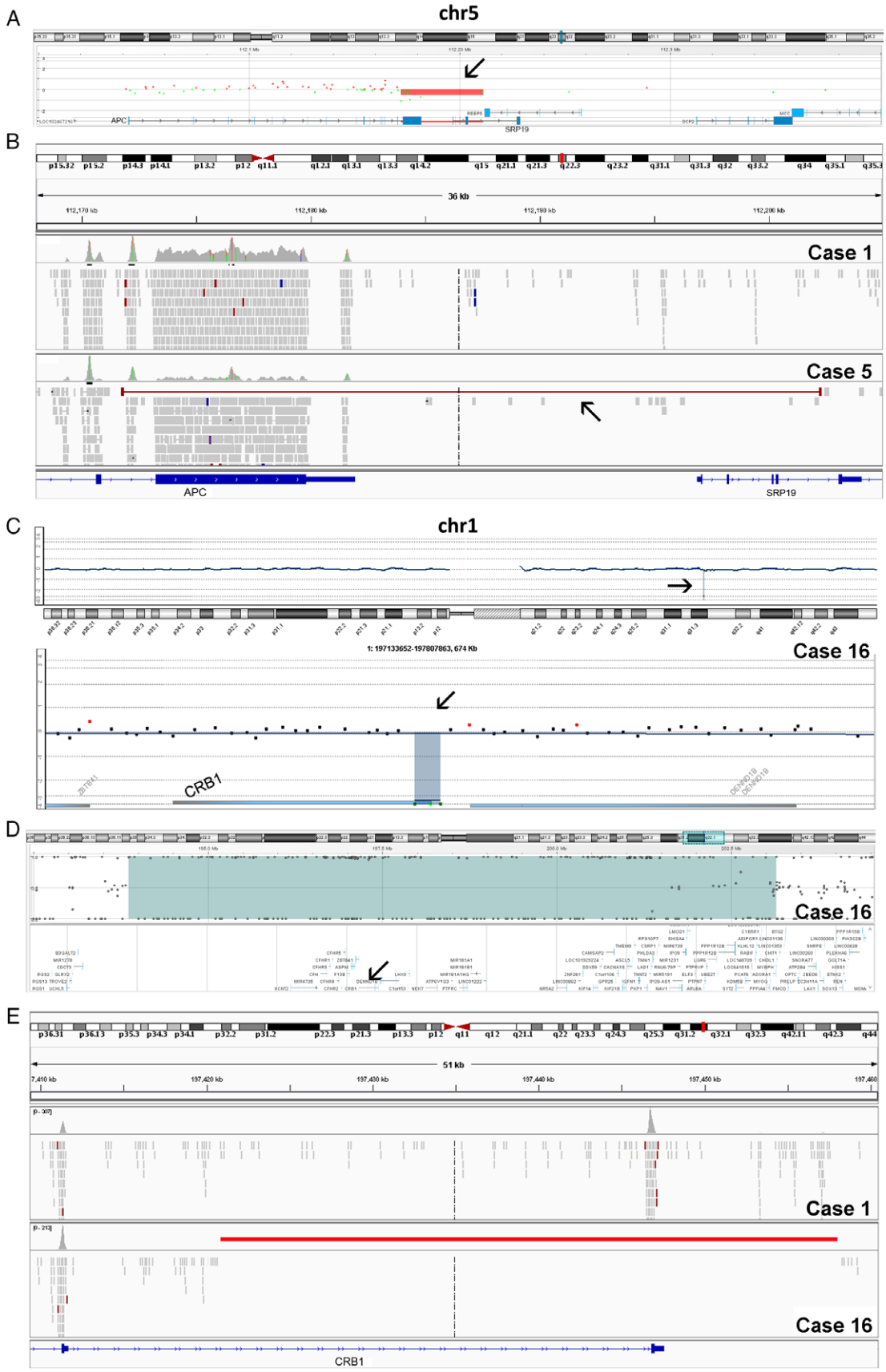

FIGURE 4 Legend on next page.

FIGURE 4 Single exon deletions: cases 5, 6 and 16. A, Screenshot of a 400 kb window from SureCall showing the approximate 30 kb deletion spanning from the intron 14 of *APC* to the intron 4 of *SRP19* (arrow); B, IGV screenshot of a 36 kb window corresponding to chr5: 112 168 000-112 205 000 in cases 1 (here used as an illustration of a normal profile in this region) and 5. In the latter a reduced coverage is noticeable, as well as the presence of a read pair (arrow) with insert size of 30 597 bp, encompassing the breakpoints of the deletion. C, Array-CGH profile of case 16 showing the homozygous 1q31.33 deletion, highlighted by arrows both in the chromosome (upper panel) and in the gene (lower panel, shaded area) views. D, Screenshot of a ~8 Mb window from SureCall showing a region of LOH (shaded area) that includes the *CRB1* gene (arrow), ascribable to the consanguinity between the parents. E, IGV screenshot of a 51 kb window corresponding to chr1: 197 410 240-197 455 400 in cases 1 (normal profile) and 16. In the latter a red line highlights the absence of reads in the region chr1: 197 420 690-197 458 350

parameters. Reads were aligned to the reference genome (GRCh37/hg19) using SureCall BWA-mem alignment tool. Post-alignment processing includes a 100 bp region padding, removal of duplicates and singletons (reads without a mate), followed by a normalization step including GC and mappability corrections. CNVs were detected comparing test sample to a sex-matched reference sample.

Detailed description of analysis processes is available in SureCall help guide. Briefly, to minimize noise due to outliers, Surecall applies a summarization method to get the central tendency of read distributions over the genomic regions covered by the OneSeq backbone. Then, the software generates $\log_2$ ratios by dividing read depth of sample over reference for each interval for which breakpoints are combined and ranked. Those intervals are candidates for amplification and deletion calls. cnLOH are determined using the SNPs covered by the backbone part of the design. SNVs and InDels are called using SureCall SNPPET algorithm and variants are annotated by using the following sources: ClinSNP_260912, ClinVarAnnotations, Cosmic_V61_260912, gwasV1_ucsc_260912 and Hs_hg19_Gene_20110426_cds.

SNVs and InDels variants were filtered according to the usual workflows,[28] considering their frequency in the general population, the indication of their disease-causing potential by in silico prediction tools, and segregation analysis in family members, either affected or healthy but potentially carriers. CNVs were evaluated for their gene content, their frequency in the general population (http://dgv.tcag.ca/dgv/app/home), and their possible overlap with those found in patients sharing comparable clinical features, according to DECIPHER (https://decipher.sanger.ac.uk) and ClinGen CNVs, (https://www.clinicalgenome.org/). Variants nomenclature was generated according to the guidelines of the Human Genome Variation Society (HGVS, http://www.hgvs.org/mutnomen) referring to the following reference transcripts: NM_005105.4 (RBM8A), NM_004070.3 (CLCNKA), NM_000085.3 (CLCNKB), NM_000492.3 (CFTR), and NM_001145853.1 (WFS1).

3 | RESULTS AND DISCUSSION

The number of sequencing reads passing mapping quality filters in BAM files ranged from 80 to 110 million for all samples. Percentage of duplicate reads was from 6% to 9%. On-target reads ±100 bp were from 67% to 75%. Average read depth in target regions was ×130, with >92% of target bases covered by 50 reads. The average number of exonic non-synonymous or splice-site variants was of 3612/sample, with an average of 159 variants/sample with MAF < 0.01, according to the available databases. These numbers were in line with those expected for a smaller exons target design with respect to a standard WES experiment (16 Mb vs approximately 50 Mb design). The average number of CNVs per sample was less than 5, comparable to that detected by 180 K array-CGH platforms. Most of the CNVs were already reported in the Database of Genomic Variants, or were inherited by a healthy parent.

In all cases except 2, both EXCAVATOR and SureCall identified the previously characterized causative genomic alteration. In contrast, the ~30 kb heterozygous deletion of case 5 was called by SureCall only, whereas the ~21 kb homozygous deletion in case 16 could be detected only by manually looking at the aligned reads. The detailed results are reported in Table 1.

3.1 | Disease-associated CNVs

In samples 1 to 4, 6 and 7, we correctly detected all the CNVs previously identified by array-CGH. These included large rearrangements visible by conventional cytogenetics and SMCs (Figure 1, cases 1, 6 and 7), and microdeletions/duplications ranging from 260 kb (case 4) to 1.4 Mb (case 2) (Figure S2A). Among them, the duplication of the non-coding *RevSex* region located upstream of the *SOX9* gene, responsible for testis development in 46,XX individuals (case 3),[16] would have been missed by conventional WES approach as well as by low resolution array platforms.

In cases 6 and 7, the tetrasomic regions (copy number >3) due to the presence of inverted duplicated SMCs[29] were correctly classified by our approach.

Cases 8 to 10 (Figure S1) had a mosaic condition consisting of a variable percentage of cells bearing trisomy for a whole chromosome (case 9) or only a portion of it (case 8 and 10). Somatic mosaicism for both structural and SNPs variants is more and more recognized as an important cause of disease. However, its detection may be problematic for several reasons such as the absence or the low grade of cells with the causative alterations in the examined tissue, which in most of the cases is blood. Crucial parameters for the detection of structural mosaicism are the size (the larger the size, the more probably the detection), the array platform (SNP- vs CGH-array), and the type of CNV (gain vs loss). The use of proper software tools for mosaic detection on approximately 1 million probes SNP chips grants a detection sensitivity, for events of at least 2 Mb size, in approximately 10% to 90% of cells for loss and LOH events, and in approximately 20% to 80% of cells for the gain ones.[30,31] In our cases, the mosaic duplication was of a size much larger than 2 Mb and was present in more than 50% of cells, as estimated by conventional cytogenetics.

3.2 | Genomic alterations associated with autosomal-recessive conditions

In 5 patients/fetuses (cases 11-15, Figures 2 and 3) affected by autosomal-recessive conditions the genomic alteration consisted of compound heterozygosity, with a point mutation in one allele and a deletion, ranging from 21 to 400 kb, in the corresponding region of the other allele. This condition has been estimated in 2.2% of cases with a recessive disease (n = 4/181), analyzed by WES[32] and in about 6% of the cystic fibrosis cases reported in HGMD, taken as an example of autosomal-recessive disease. Further, in a patient (case 17, Figures S2B, C), maternal isodisomy for chromosome 4 reduced to homozygosity a heterozygous variant in the *WFS1* gene, c.1348_1350delinsTAG.[22] Uniparental disomy was estimated to account for 2.8% (n = 5/181) of homozygous recessive diseases in the Yang *et al.* cohort.[32] Studies by SNP-arrays estimated that whole-chromosome isodisomy is very rare and that in most of the cases the reduction to homozygosity in a single chromosome is of the segmental type, with isodisomy regions interspersed with others in heterodisomy.[6]

These cases were probably those most benefited by the OneSeq assays, allowing to get the final molecular diagnosis in a single experiment. In fact, in all these individuals the conclusive molecular diagnosis was only reached after several tests, including Sanger sequencing, MLPA and genome-wide array analysis.

3.3 | Critical cases

The deletion in cases 5 and 16 was not detected by EXCAVATOR under standard parameters but we were anyway able to pick-up and characterize them at bp level, by SureCall (case 5) and/or manual inspection of the reads (case 16) (Figure 4).

In case 5, a 30.4 kb heterozygous deletion spanning from the last intron of *APC* to the intron 4 of *SRP19* was originally detected by array-CGH and the breakpoints mapped by Sanger sequencing.[17] Since *SRP19* is not a disease-associated gene, it is not enriched at exon level in the OneSeq panel and the reads covering the last *APC* exon were probably insufficient to provide enough coverage to correctly detect the deletion. Thus the CNV was not called at all by EXCAVATOR, whereas SureCall could identify it, although by a $\log_2$-ratio of −0.5, lower than expected one (−1). By manually inspecting the reads mapping to chr5: 112 168 000-112 205 000 (hg19), a reduced coverage of the deleted region was evident with respect to the rest of the gene and to the other samples of the same experiment (Figure 4A,B). Moreover, we identified a single read pair with insert size of 30 597 bp, encompassing the breakpoints of the deletion (Figure 4B, arrow), thus providing the junction breakpoint position at base-pair resolution (chr5: 112 171 792-112 202 336). This last observation suggests the need for further bioinformatics tools leveraging paired-read and/or split-reads information to detect gross insertion and deletions.[33,34]

In case 16, a 21.2 kb homozygous deletion including the last exon of *CRB1* was identified by a 180 k array-CGH platform: arr [hg19] 1q31.33(197 434 389-197 455 060)x0 pat mat (Figure 4C). Both parents, coming from the same village of Sardinia, were heterozygous carriers of the deletion.

Although SureCall indicated the presence of a large LOH region (approximately 8 Mb) surrounding *CRB1* (Figure 4D), the deletion was not called. Manual inspection showed a complete lack of reads mapping in the region chr1: 197 420 690-197 458 350, including the off-target ones (Figure 4E). In this case, no reads mapping on the breakpoints were available.

4 | CONCLUSIONS

The purpose of our study was to investigate the ability of a single test, based on target enrichment and NGS, to identify different types of genomic lesions, including SNVs/InDels, LOH, and large deletions/duplications not necessarily containing protein-coding genes.

Different approaches for CNVs detection from WES data have already been reported, demonstrating an increased diagnostic yield up to 6% (an average of 2%), that can be obtained without additional direct laboratory costs, but by optimization of the data analysis.[35]

In our cohort, 3 cases (cases 1, 11 and 12) were fetuses with severe ultrasound abnormalities requiring invasive sampling, cytogenetic and molecular investigations. Moreover, the condition affecting other 7 cases might have been suspected prenatally (cases 2, 3, 7, 9, 13-15). In all these cases a targeted sequencing approach integrating SNVs/InDels, CNVs and LOH analysis in a single NGS experiment, would have had the undoubted advantage of a rapid diagnosis. These findings further strengthen that our approach is particularly useful in a prenatal setting, where the time-frame for a genetic diagnosis is short.

In our cohort, the turnaround time from DNA extraction to data analysis and reporting was on average 15 days that could be possibly reduced by the novel approaches for faster library preparation. Since the platform we used contains only genes whose variants associate with known pathological conditions, the need for extended segregation studies in the family or in other families with similar pathologies is negligible. This fact dramatically shortens investigation times over the entire exome platforms, especially if parental and sibship DNA is already available for Sanger analysis.

Genomic arrays are able to diagnose up to 10% of fetal malformations with normal karyotype,[36,37] whereas the application of WES, with or without CNV detection, has demonstrated a variable percentage of successful diagnosis in small cohorts of fetuses with ultrasound abnormalities.[38,39]

This study demonstrates that OneSeq was able to detect a wide range of genomic events, largely overcoming the limitations for CNVs detection by WES, mainly due to the uneven distribution of the reads, restricted to exons. We obtained high-level concordance between the 2 different pipelines we used, and between the size of the CNVs identified by the array and OneSeq approach (Table 1). Data analysis remains the major bottleneck of NGS in general and CNV detection particularly.

As for the minimal resolution for CNVs analysis in diagnostics, the panel we used, having a backbone resolution of 300 kb, is able to detect CNVs of 400 kb or larger, as recommended by the ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis.[40]

This approach is a good compromise as long as WGS will be available in diagnostics, thanks to lower costs and proper interpretation algorithms. Software development and standardization as well as large prospective cohort studies are required to reinforce the benefit of every possible panel allowing combined detection of CNV, SNV and cnLOH.

ACKNOWLEDGEMENTS

We acknowledge Prof. Mariano Rocchi and Dr. Oronzo Capozzi, University of Bari, for their contribution to FISH analyses. A.V. benefits of a research position granted by the University of Pavia in the context of the strategic plan: "Towards a governance model for the international migration: an interdisciplinary and diachronic perspective." O.Z. is supported by Telethon Italy, Grant: GGP13060.

Conflict of interest

All authors, except for D.G. and I.L. declare no conflict of interest. D.-G. and I.L. are employed at Agilent Technologies. A.V. and O.-Z. received part of the consumables by Agilent Technologies. The study was approved by the local Ethics Committee of Meyer Children's University Hospital. Informed written consent was obtained from the parents or legal guardians of each participant. The DNA samples were anonymized by using alphanumeric code.

REFERENCES

1. Gazzo AM, Daneels D, Cilia E, et al. DIDA: a curated and annotated digenic diseases database. *Nucleic Acids Res*. 2016;44:D900-D907.
2. Schaffer AA. Digenic inheritance in medical genetics. *J Med Genet*. 2013;50:641-652.
3. 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.
4. Rovelet-Lecrux A, Hannequin D, Raux G, et al. APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat Genet*. 2006;38:24-26.
5. Van Esch H. MECP2 Duplication Syndrome. In: Pagon RA, Adam MP, Ardinger HH, et al., eds. *GeneReviews*. Seattle, WA: University of Washington, Seattle; 1993-2017.
6. King DA, Fitzgerald TW, Miller R, et al. A novel method for detecting uniparental disomy from trio genotypes identifies a significant excess in children with developmental disorders. *Genome Res*. 2014;24:673-687.
7. Wiszniewska J, Bi W, Shaw C, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur J Hum Genet*. 2014;22:79-87.
8. Sawyer SL, Hartley T, Dyment DA, et al. Utility of whole-exome sequencing for those near the end of the diagnostic odyssey: time to address gaps in care. *Clin Genet*. 2016;89:275-284.
9. Stenson PD, Mort M, Ball EV, et al. The human gene mutation database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet*. 2014;133:1-9.
10. Albers CA, Paul DS, Schulze H, et al. Compound inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet*. 2012;44:435-439.
11. Boone PM, Campbell IM, Baggett BC, et al. Deletions of recessive disease genes: CNV contribution to carrier states and disease-causing alleles. *Genome Res*. 2013;23:1383-1394.
12. de Ligt J, Boone PM, Pfundt R, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. *Hum Mutat*. 2013;34:1439-1448.
13. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res*. 2012;22:1525-1532.
14. Hehir-Kwa JY, Pfundt R, Veltman JA. Exome sequencing and whole genome sequencing for the detection of copy number variation. *Expert Rev Mol Diagn*. 2015;15:1023-1032.
15. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet*. 2015;24:R102-R110.
16. Vetro A, Dehghani MR, Kraoua L, et al. Testis development in the absence of SRY: chromosomal rearrangements at SOX9 and SOX3. *Eur J Hum Genet*. 2015;23:1025-1032.
17. Quadri M, Vetro A, Gismondi V, et al. APC rearrangements in familial adenomatous polyposis: heterogeneity of deletion lengths and breakpoint sequences underlies similar phenotypes. *Fam Cancer*. 2015;14:41-49.
18. Forlino A, Vetro A, Garavelli L, et al. PRKACB and Carney complex. *N Engl J Med*. 2014;370:1065-1067.
19. Papoulidis I, Oikonomidou E, Orru S, et al. Prenatal detection of TAR syndrome in a fetus with compound inheritance of an RBM8A SNP and a 334 kb deletion: a case report. *Mol Med Rep*. 2014;9:163-165.
20. Corbetta S, Raimondo F, Tedeschi S, et al. Urinary exosomes in the diagnosis of Gitelman and Bartter syndromes. *Nephrol Dial Transplant*. 2015;30:621-630.
21. Costantino L, Rusconi D, Claut L, et al. A wide methodological approach to identify a large duplication in CFTR gene in a CF patient uncharacterised by sequencing analysis. *J Cyst Fibros*. 2011;10:412-417.
22. Papadimitriou DT, Manolakos E, Bothou C, et al. Maternal uniparental disomy of chromosome 4 and homozygous novel mutation in the WFS1 gene in a paediatric patient with Wolfram syndrome. *Diabetes Metab*. 2015;41:433-435.
23. Vetro A, Manolakos E, Petersen MB, et al. Unexpected results in the constitution of small supernumerary marker chromosomes. *Eur J Med Genet*. 2012;55:185-190.
24. Rehm HL, Berg JS, Brooks LD, et al. ClinGen–the clinical genome resource. *N Engl J Med*. 2015;372:2235-2242.
25. Magi A, Tattini L, Cifola I, et al. EXCAVATOR: detecting copy number variants from whole-exome sequencing data. *Genome Biol*. 2013;14:R120.
26. Vetro A, Savasta S, Russo Raucci A, et al. MCM5: a new actor in the link between DNA replication and Meier-Gorlin syndrome. *Eur J Hum Genet*. 2017;25:646-650.
27. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178-192.
28. Seaby EG, Pengelly RJ, Ennis S. Exome sequencing explained: a practical guide to its clinical application. *Brief Funct Genomics*. 2016;15:374-384.
29. Amor DJ, Choo KH. Neocentromeres: role in human disease, evolution, and centromere study. *Am J Hum Genet*. 2002;71:695-714.
30. Jacobs KB, Yeager M, Zhou W, et al. Detectable clonal mosaicism and its relationship to aging and cancer. *Nat Genet*. 2012;44:651-658.
31. King DA, Jones WD, Crow YJ, et al. Mosaic structural variation in children with developmental disorders. *Hum Mol Genet*. 2015;24:2733-2745.
32. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312:1870-1879.
33. Layer RM, Chiang C, Quinlan AR, et al. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol*. 2014;15:R84.
34. Rausch T, Zichner T, Schlattl A, et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28:i333-i339.
35. Pfundt R, Del Rosario M, Vissers LE, et al. Detection of clinically relevant copy-number variants by exome sequencing in a large cohort of genetic disorders. *Genet Med*. 2017;19:667-675.

36. Hillman SC, McMullan DJ, Hall G, et al. Use of prenatal chromosomal microarray: prospective cohort study and systematic review and meta-analysis. *Ultrasound Obstet Gynecol*. 2013;41:610-620.
37. Srebniak MI, Diderich KE, Joosten M, et al. Prenatal SNP array testing in 1000 fetuses with ultrasound anomalies: causative, unexpected and susceptibility CNVs. *Eur J Hum Genet*. 2016;24:645-651.
38. Carss KJ, Hillman SC, Parthiban V, et al. Exome sequencing improves genetic diagnosis of structural fetal abnormalities revealed by ultrasound. *Hum Mol Genet*. 2014;23:3269-3277.
39. Drury S, Williams H, Trump N, et al. Exome sequencing for prenatal diagnosis of fetuses with sonographic abnormalities. *Prenat Diagn*. 2015;35:1010-1017.
40. South ST, Lee C, Lamb AN, et al. ACMG standards and guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet Med*. 2013;15:901-909.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Vetro A, Goidin D, Lesende I, et al. Diagnostic application of a capture based NGS test for the concurrent detection of variants in sequence and copy number as well as LOH, . *Clin Genet*. 2017;1–12. https://doi.org/10.1111/cge.13060