

ARTICLE

Discovery of variants unmasked by hemizygous deletions

Ron Hochstenbach^{1,3}, Martin Poot^{*,1,3}, Isaac J Nijman^{2,3}, Ivo Renkens¹, Karen J Duran¹, Ruben van't Slot¹, Ellen van Binsbergen¹, Bert van der Zwaag¹, Maartje J Vogel¹, Paulien A Terhal¹, Hans Kristian Ploos van Amstel¹, Wigard P Kloosterman¹ and Edwin Cuppen^{1,2}

Array-based genome-wide segmental aneuploidy screening detects both *de novo* and inherited copy number variations (CNVs). In sporadic patients *de novo* CNVs are interpreted as potentially pathogenic. However, a deletion, transmitted from a healthy parent, may be pathogenic if it overlaps with a mutated second allele inherited from the other healthy parent. To detect such events, we performed multiplex enrichment and next-generation sequencing of the entire coding sequence of all genes within unique hemizygous deletion regions in 20 patients (1.53 Mb capture footprint). Out of the detected 703 non-synonymous single-nucleotide variants (SNVs), 8 represented variants being unmasked by a hemizygous deletion. Although evaluation of inheritance patterns, Grantham matrix scores, evolutionary conservation and bioinformatic predictions did not consistently indicate pathogenicity of these variants, no definitive conclusions can be drawn without functional validation. However, in one patient with severe mental retardation, lack of speech, microcephaly, cheilognathopalatoschisis and bilateral hearing loss, we discovered a second smaller deletion, inherited from the other healthy parent, resulting in loss of both alleles of the highly conserved heat shock factor binding protein 1 (*HSBP1*) gene. Conceivably, inherited deletions may unmask rare pathogenic variants that may exert a phenotypic impact through a recessive mode of gene action.

European Journal of Human Genetics (2012) 20, 748–753; doi:10.1038/ejhg.2011.263; published online 18 January 2012

Keywords: idiopathic mental retardation; inherited deletion; mutation unmasking; expressive language disorder; *HSBP1*

INTRODUCTION

Application of array-based genome-wide copy number investigation to patients with idiopathic mental retardation or developmental delay (MR/DD), with or without multiple congenital abnormalities (MCA), has identified pathogenic segmental aneuploidies in up to 19% of consecutive referrals.^{1–3} *De novo* deletions in sporadic patients are generally believed to be pathogenic. In contrast, genes within an inherited deletion may not provoke a phenotypic effect by mere haploinsufficiency, as these genes were also found in a single copy in a healthy parent. However, such inherited deletions may still contribute to disease when combined with rare *de novo* or inherited damaging alleles on the other chromosome. Therefore, screening for variants in regions corresponding to inherited hemizygous deletions may allow us to identify genes that exert a phenotypic effect by independent loss of both alleles.

Assuming this recessive mode of inheritance, genes within a deletion from a healthy parent may be pathogenic if a second mutation within the deleted region has been transmitted by the other, equally healthy parent or has arisen *de novo*. This recessive mechanism was initially proposed by Hatchwell.⁴ A literature survey shows that in multiple cases pathogenic single-nucleotide variants (SNVs) combined with a *de novo* deletion result in aberrant phenotypes (Supplementary Table 1). However, only few cases demonstrated unmasking by an inherited deletion (Supplementary Table 2). Whether 'unmasking heterozygosity' is a frequent mechanism for abnormality' is being debated.^{5,6} The recent flurry of publications

regarding losses of a single or several exons in selected genes by using targeted arrays underscore that current routine diagnostic methods may miss potentially pathogenic copy number variations (CNVs).^{7–15}

In this study, we investigated whether unmasking of recessive alleles may contribute to the phenotype of patients with idiopathic MR/MCA by analyzing all coding sequences in hemizygous deletion regions of patients who inherited a novel deletion from a healthy parent. To do so, we performed multiplexed genomic enrichment and next-generation sequencing of the entire coding sequence of all genes in the deletion regions of 20 patients. We identified eight SNVs that were located exclusively within the corresponding deletions of patients. In one of the patients the inherited deletion unmasked another small-inherited deletion on the second allele, which comprised the heat shock factor binding protein 1 (*HSBP1*) gene. Taken together, this approach provides a sensitive method to detect mutations on the second chromosome within hemizygous deletions.

MATERIALS AND METHODS

The procedures for selection of patients with an inherited deletion and the molecular investigations performed^{16,17} are described in Supplementary Note 1 and Supplementary Tables 3 and 4.

Enrichment array design

For each patient, the maximum deletion was defined as the interval between the non-deleted array-probes flanking the deletion. When a gene was interrupted

¹Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; ²Hubrecht Institute and University Medical Center Utrecht, Utrecht, The Netherlands

*Correspondence: Dr M Poot, Department of Medical Genetics, University Medical Centre Utrecht, PO Box 85090, 3508 AB Utrecht, The Netherlands. Tel: +31 88 75 53800; Fax: +31 88 75 53801; E-mail: m.poot@umcutrecht.nl

³These authors contributed equally to this work.

Received 18 August 2011; revised 13 December 2011; accepted 14 December 2011; published online 18 January 2012

by a breakpoint, all exons of that gene were included. Using the strategy developed by Mokry *et al.*¹⁸ we designed arrays of 60-mer probes tiled with an off-set of 10 nt covering the entire coding sequence of all genes in the deletion regions according to the Ensembl database (www.ensembl.org) using the GRCh37/hg19 build of the human reference genome. One capture array was designed for the deletion intervals of patients 1, 9, 10, 12, 13, 14, 18, 19, 20 and a second capture array was designed for patients 2, 3, 4, 5, 6, 7, 8, 11, 15, 16 and 17. Probes were synthesized on 1 M custom Agilent arrays.

Library preparation, enrichment and massively parallel DNA sequencing

For each patient, 1 μ g genomic DNA was used as input for DNA shearing to fragments of 100–120 nt in length. Library preparation was performed as described by Mokry *et al.*¹⁸ After ligation of short adaptors (adaptor 1 and adaptor 2, Supplementary Table 4), SOLiD library barcodes (barcodes 1–20, SOLiDv3.5 manual, Life Technologies, Carlsbad, CA, USA) were introduced by 5–8 PCR cycles (100 ng adaptor-ligated DNA, 1 ml Platinum PCR Supermix (Invitrogen, Carlsbad, CA, USA), 1 μ l Pfu Turbo DNA polymerase (Promega, Madison, WI, USA), 6 μ l of short-P1-primer (50 μ M) and 6 μ l of barcode primer 1–20 (50 μ M). Libraries were pooled after size selection (one pool for each array) and 3 μ g pooled DNA was mixed with 15 μ g C₀T-1 DNA (1 mg/ml) and 1.5 μ l of each barcode blocker oligo (barcode blocker 1 and 2, 10 μ g/ μ l each) and hybridized on Agilent capture arrays as described by Mokry *et al.*¹⁸ Post enrichment PCR was performed using 1 ml Platinum PCR Supermix, 1 μ l Pfu Turbo DNA polymerase, 6 μ l of P1-primer and 6 μ l of P2-primer for 12 cycles. Enriched libraries were sequenced on a single slide of a SOLiDv3.5 sequencer following standard procedures (Supplementary Table 5).

DNA sequence analysis and mutation identification

Sequencing reads were mapped against the reference genome (GRCh37, ensembl59) using the Burrows-Wheeler Alignment Tool¹⁹ and all SNVs and short indels up to 7 bp were compiled and annotated using in-house developed single-nucleotide polymorphism-calling (SNP) algorithms using the ensemble API release 59.²⁰ In this compilation, SNVs mapping within the deletion region of each patient were selected for confirmation by capillary sequencing, following standard protocols. First, we compiled for each DNA sample all SNVs with a non-reference allele call >90%. Second, we compared SNVs between DNA samples, and third, we retained those SNVs, which occurred only once in the samples studied and only within the cognate deletion region of each patient. Confirmed SNVs were subsequently traced in the DNA of the parents. In addition, by capillary sequencing loss of exon 3 of the *HSBP1* gene was confirmed in the affected child. The primer pairs used for confirmation by capillary sequencing are listed in Supplementary Table 6. All confirmed SNVs were evaluated for their phenotypic impact using the Grantham matrix score,²¹ Genomic Evolutionary Rate Profile (GERP),^{22,23} PANTHER²⁴ and PolyPhen2.²⁵ European population frequencies were given for variants with data other than 1000 genome pilot study. We used bioinformatics tools to identify copy number changes based on read coverage deviations from normalized control levels by normalizing the coverage level of each nucleotide for the haploid patient DNA versus the mean coverage level of the other 8 or 10 diploid DNA samples in the same pool by a custom script written in PERL. The detected regions with significant (more than two standard deviations from the mean) coverage abnormalities were supported by DWAC-Seq (<http://dwacseq.hubrecht.eu>).

RESULTS

Detection of SNVs in hemizygous deletions

From a cohort of ~1500 patients with MR/MCA we selected 19 patients with a novel hemizygous deletion inherited from a healthy parent and 1 patient who had a *de novo* hemizygous deletion (Supplementary Table 3). All 20 DNA samples were, in two pools of 9 and 11 samples, respectively, subjected to multiplexed enrichment and SOLiDv3.5 sequencing to an average coverage depth of 256 \times (Supplementary Table 5). Sequence reads were mapped to the human

reference genome (GRCh37/hg19) and 703 SNVs with a predicted effect at the protein level were identified in the individual samples with in-house developed SNV-calling software.²⁰ In addition we found 36 indels of maximally 7 bp, none of which confirmed to the criteria of being potentially pathogenic as described above (Materials and Methods). For three samples the SOLiD-based SNVs were compared with SNP data generated with Illumina Infinium HumanHap300 Genotyping BeadChips. For 240 (98.4%) out of the 244 SNPs covered by both platforms full concordance was found, while for 4 SNPs (1.6%) the platforms disagreed. For these four SNPs, both platforms disagreed for these being either homozygous non-reference or heterozygous. In total, the SOLiD-based data contained 675 non-synonymous SNVs calls, 15 stops gained and 13 stops lost (GRCh37/hg19); 642 SNVs were listed in dbSNP build 131 and 61 SNVs were novel. All stops gained and lost were listed as known polymorphisms in dbSNP and none of the variants were listed in databases of known pathogenic variation (eg, HGMD).

In seven patients we found eight non-synonymous SNVs mapping exclusively to the cognate deletion regions, which we subsequently confirmed by capillary sequencing (Table 1). These SNVs showed clearly homozygous calls and seven were known SNPs (according to dbSNP build 131). To assess whether the known SNPs were likely to be compatible with healthy life we determined the frequencies of homozygous calls of this particular allele in the European population as recorded in dbSNP. For three SNPs (rs5743820, rs2227983 and rs4149056) such data were available; they indicate that these SNPs have been found at frequencies ranging from 0.0016 to 0.2649 for the homozygous state in the healthy European population. These limited data are consistent with these SNPs being in Hardy-Weinberg equilibrium. Of note is that none of the genes, in which these SNPs were found, has previously been found mutated in patients with MR/MCA (see Supplementary Table 4). Interestingly, a novel SNP in the *FAM178B* gene in a severely dyslexic patient (patient no. 13) with a deletion of this gene inherited from his mother, resulted in the same genotype in the patient at this particular position in the *FAM178B* gene as that of his language proficient mother. This rules out a pathogenic contribution of this hemizygous C allele in the causation of his dyslexia.

To assess whether these SNVs may contribute to a clinical phenotype, we determined their putative impact upon protein structure and function as reflected by the Grantham matrix scores.²¹ All SNVs had Grantham matrix scores in the low range, suggesting an only minor impact. To assess the degree of evolutionary conservation of a SNV we determined the GERP score.^{22,23} Negative GERP scores reflect no evolutionary conservation, whereas positive scores are in agreement with a pathogenic impact of a SNV.^{22,23} All SNVs, except those in the *ZNF800* and the *OSGIN1* genes showed low-positive GERP scores, suggesting little likelihood of pathogenic impact. In contrast, PolyPhen2,²⁵ which combines the impact on protein structure and function with the level of evolutionary conservation, predicted a highly damaging effect from the T to C transition of rs4149056 in the *SLCOB1* gene, while PANTHER²⁴ returned no score and was thus non-informative. Thus, this SNV with a Grantham Matrix Score of 64, a GERP score of -1.19 and occurring in the homozygous state in the European population at 10.66% was judged 'damaging' with a likelihood of 0.993 by PolyPhen2 (Table 1). In summary, currently available algorithms predicting impact on protein structure and function, determining evolutionary conservation and combining multiple parameters did not return consistent predictions regarding pathogenicity of the eight non-synonymous SNVs in our data set.

Table 1 Genotypes, inheritance, Grantham matrix score, GERP, PANTHER and PolyPhen2 evaluation of all SNVs located exclusively to the cognate deletion regions of the patients

Patient no.	Coordinate ^a	Gene name	Genotype patient	Genotype Father	Genotype Mother	HGVS ^b	% non-ref ^c	Coverage ^d	dbSNP ID (v1.31)	Frequency (EUR)	Grantham ^e	GERP ^f	PANTHER ^g	PolyPhen2 ^h (probability)
1	4:g38828288G>A	TLR6	A/del	G/del	GA	NM_006068.4 c.2267C>T (p.Thr756Met)	99	760	rs5743820	0.0016	81	2.4	0.2708	Benign (0.076)
2	7:g54621692G>A	VSTM2A	A/del	AA	G/del	ENSTR00000404951 c.767G>A (p.Arg256Gln)	100	162	rs68106607	?	43	-5.09	?	?
2	7:g5522925G>A	EGFR	G/del	GG	A/del	NM_005228.3 c.1562G>A (p.Arg521Lys)	100	86	rs2227983	0.2649	26	-6.21	?	Benign (0.000)
4	7:g127015083G>A	ZNF800	A/del	GA	G/del	NM_176814.3 c.307C>T (p.Pro103Ser)	100	72	rs62621812	?	74	4.13	0.1315	Benign (0.009)
8	12:gg21331549T>C	SLCO1B1	C/del	T/del	TC	NM_006446.4 c.521T>C (p.Val174Ala)	100	185	rs4149056	0.1066	64	-1.19	?	Prob. dam (0.957)
9	2:g54035508G>C	ERLEC1	C/del	CC	G/del	NM_015701.3 c.952G>C (p.Val318Leu)	100	391	rs2287345	?	32	-8.85	0.2963	Benign (0.000)
13 ^h	2:gg76333267G>C	FAM178B	C/del	GC	C/del	CDS46366 c.628C>G (p.Pro210Ala)	100	155	rs113319086	?	27	2.99	0.1843	Prob. dam (0.980)
15	16:gg8399895G>A	OSGNI	A/del	GA	G/del	NM_013370.3 c.1066G>A (p.Glu356Lys)	100	133	novel	?	56	3.55	0.1885	Prob. dam. (0.966)

Abbreviations: GERP, genomic evolutionary rate profile; SNP, single-nucleotide polymorphism; SNV, single-nucleotide variants.

^aGenome coordinates are according to the Human Reference Sequence GRCh37/hg19 (February 2009). Genotypes were confirmed by capillary sequencing.

^bAmino-acid change in HGVS notation. A representative transcript is shown.

^cPercentage of sequence reads in the patient sample, depicting a nucleotide being not the one in the human reference genome.

^dNumber of sequence reads in the patient sample covering the SNV.

^eThe Grantham matrix score gives the impact of the amino acid change on protein structure and function (1.7), a score below 100 indicates low impact; a score > 100 indicates significant impact.

^fGERP refers to Genomic Evolutionary Rate Profile based on a 13 species comparison;^{18,19} Scores < 3.0 indicate little evolutionary conservation; scores > 3.0 indicate strong evolutionary conservation.

^gPANTHER²¹ and PolyPhen2²⁰ return probabilities of a SNV being deleterious based on weighted consideration of multiple functional and evolutionary parameters. HumDiv value given.

^hThis severely dyslexic patient has the same genotype at the FAM178B locus as language proficient mother, thereby ruling out a pathogenic contribution of this hemizygous C allele in the causation of his dyslexia. ? indicates no data available.

Detection of hemizygous deletions by coverage depth analysis

Subsequently, we used bioinformatics tools to identify copy number changes based on read coverage deviations from normalized control levels.²⁶ Thus, plots of the deletion regions flanked by diploid regions were obtained (Supplementary Figure 1). Although 11 DNA samples showed approximately half the coverage depth relative to the other samples from the same pool for the deletion region, 9 samples showed erratic patterns. Coverage signals approaching the diploid level within deletion areas corresponded with segmental duplications or indels (results not shown). Segmental duplications and indels were not detected by BAC-, oligonucleotide and SNP arrays, as these regions were not covered by these platforms. As these signals locate in regions of sequence homology with a locus outside the targeted deletion region, they may reflect more complex haplotypes in our patients than represented in the human reference genome.²⁶

Unmasking of an inherited deletion of the HSBP1 gene

The deletion region of patient 15, with a normalized coverage level almost entirely half of that of the flanking regions, showed a small region with close to zero coverage (Figure 1a). This region corresponded to three probes with lowered signals on the oligonucleotide array-CGH (Figure 1b) and contained the entire HSBP1 gene (Figure 1c). PCR-based resequencing of exon 3 of HSBP1 showed complete absence in the affected child, with presence of the normal sized PCR product, indicative of at least one intact copy, in both parents (results not shown). Comparing the oligonucleotide array and the SOLiD coverage depth data the proximal breakpoint of this deletion maps to nucleotide position 83 841 341 (SOLiD) and the distal breakpoint between 83 857 382 and 83 873 030 (oligonucleotide array data). The latter breakpoint could not be mapped more precisely, because of the absence of resequenced exonic sequences in this region. This combined data indicate a bi-allelic deletion of minimally 16 041 and maximally 31 689 bp, which encompasses HSBP1, but no other gene (Figure 1c). We conclude that both alleles of the entire HSBP1 gene have been lost in the patient. As both heterozygous parents were healthy, the absence of both gene copies in the child may be pathogenic when a recessive mode of gene action is assumed.

DISCUSSION

Array-based segmental aneuploidy detection has become the first-tier genetic test in MR/MCA patients.^{1,27} Thus, familial deletions are increasingly being identified in healthy probands, thereby identifying genes that seem refractory to haploinsufficiency. Several mechanisms have been put forward to explain the phenotypic differences between healthy carriers of deletions and patients within a family.^{28–30} To investigate unmasking of recessive alleles in MR/MCA we analyzed all coding sequences in hemizygous deletion regions of 20 patients using array-based multiplexed enrichment followed by next-generation sequencing. This focused approach allows the simultaneous identification of SNVs and CNVs in a DNA sample at nucleotide resolution. Furthermore, the multiplexed design of this experiment automatically generates proper controls to efficiently filter out noise and coverage level differences generated by next-generation sequencing.

After evaluation of their inheritance patterns, impact on protein structure and function and evolutionary conservation the eight SNVs exclusively found in the corresponding deletion regions of the patients did not suggest consistent and clear pathogenic effects (Table 1). This is in line with results of the sequence analysis in inherited deletions in other studies, including the 1q21.1 deletion associated with TAR (thrombocytopenia-absent radius) syndrome,^{31,32} the 16p13.11

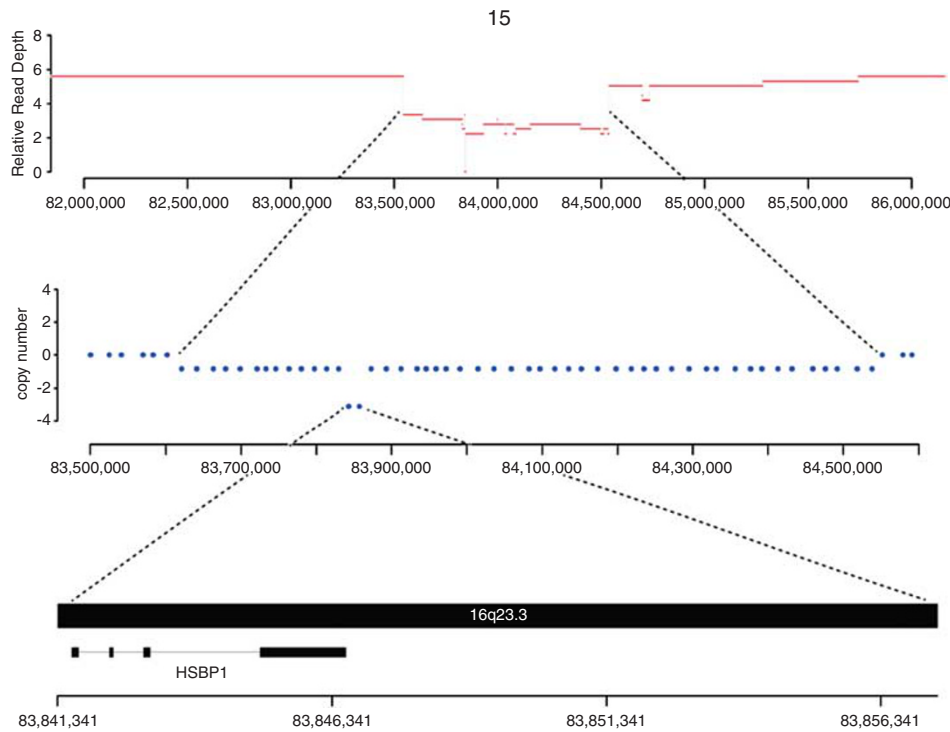


Figure 1 Oligonucleotide array and SOLiD sequencing data for the deletion region between nucleotide position 83 841 341 and 83 857 382 in chromosome band 16q23.3 in patient no. 15 (**Supplementary Table 3**). Upper panel: Normalized coverage depth for the deletion region. Relative read depths are in arbitrary units as produced as output by DWAC-seq.²⁶ Note the lower read depth for the entire deletion region and the lowered signal in the left part of this region. Middle panel: Oligonucleotide array data. Note the lowered signal for three probes in the left half of the deletion region. Bottom panel: The genome region for which both alleles have been lost in the patient (UCSC browser GRCh37/hg19). Dashed lines demarcate the corresponding deletion boundaries.

deletion, which is associated with mental retardation and autism,³³ the 16p11.2 deletion associated with autism³⁴ and the 16p12.1 deletion associated with childhood DD,³⁵ which all failed to detect pathogenic SNVs. In contrast to our approach, these studies did not include all coding regions in the deletion, but only selected candidate genes.

The single unmasked structural variation identified in our study, is much more likely to contribute to the patient's phenotype, including mental retardation without speech, optic nerve hypoplasia, nasolacrimal duct obstruction, cheilognathopalatoschisis, microcephaly and bilateral hearing loss. In DECIPHER³⁶ a single patient with a *de novo* hemizygous deletion encompassing, *HSBP1* (DECIPHER patient no. 2801) presented with a less complex phenotype. Interestingly, the hemizygous parents of our patient appeared normal. This constellation of homozygous losses in a patient with healthy parents carrying heterozygous losses is similar to that of two patients reported in the literature.³⁷

By interaction with *HSP1* (heat shock factor 1, the major transcriptional activator of heat shock genes), *HSBP1* has been shown to negatively regulate the heat shock response in *C. elegans*, in *Xenopus* tadpoles and in cultures of mammalian cells.^{38,39} Knockdown of *Hsbp1*, which is a part of the AKT1–DNA transcription network, in C57Bl/6 mice diminished neuroblast migration.⁴⁰ Involvement of *HSBP1* in brain development in mice is consistent with a phenotypic impact of nullizygosity for *HSBP1* entailing a form of mental retardation and expressive speech disorder as found in our patient. Since to this date, *HSBP1* has not been associated with the genetic pathways governing classical speech and language disorders,⁴¹ further investigations into its function appear needed.

Population analyses of large CNVs revealed a frequency of 5–10% in healthy individuals for CNVs of 500 kb and larger, which is in the

same range as in patients with DD.⁴³ As size and gene content of CNVs are negatively correlated with their population frequency,⁴² smaller structural variants may occur at even higher frequencies in the healthy population. A preliminary assessment indicates that CNVs of 500 bp and larger occur at a median frequency of more than 1000 CNVs per healthy individual.^{42,44} Intragenic deletions, encompassing one to several exons, account for 2–5% of the spectrum of deleterious mutations in Mendelian disorders.^{7–15} In case an inherited or *de novo* intragenic deletion is combined with a deletion of the entire gene transmitted by a healthy parent no intact alleles of a particular gene will be found in the patient, which amounts to a recessive mode of gene action. In agreement with this inference is the finding that numerous patients within the MR/MCA spectrum carry novel or rare CNVs at two or even more loci.^{16,17,36} The latter may further increase the phenotypic variability among patients with disorders resulting from structural genome variations.³⁶ Of note is that in contrast to the patients selected for this study, most reports of unmasking hemizygosity in the literature describe patients who combine two distinct, recognizable syndromes or diseases, such as Prader Willi Syndrome + albinism,⁴⁵ Angelman Syndrome + albinism,⁴⁶ Smith Magenis Syndrome + sensorineural hearing loss,⁴⁷ or Wolf–Hirschhorn syndrome + Wolfram syndrome⁴⁸ (see also Supplementary Tables 1 and 2), which may relate to the frequency of the mechanism under study here.

Our study of patients with idiopathic MR/MCA did not reveal any compelling evidence, except for the single-gene deletion, for unmasking of obvious gene mutations by inherited hemizygous deletions. This may be due to either the incorrect assessment of candidate genes and/or evaluation of the deleterious impact of the mutations identified in these genes by current prediction algorithms. Furthermore, we assessed only the protein-coding regions and the beginning and ends

of introns in the deletion regions and may thus have missed variants that affect functions in the UTRs of the genes, non-coding RNAs, regulatory elements, or splicing efficiency.⁴⁹ In addition, biases introduced by the enrichment technique as well as mapping of short (50 nt) next-generation sequencing reads result in an uneven coverage of the target regions. Therefore, not every coding base is assayed equally efficiently and some coding variants may thus have been missed. However, our data show that this involves less than 5% of the coding sequences (results not shown). Finally, uniparental disomies and segmental homozygosity of chromosomes in genetically isolated populations,^{50–53} mutations at an unrelated locus (see patient no. 18; Materials and Methods) and differences in imprinting levels may also come into play. These possible mechanisms may represent targets for future studies.

Although our analysis of unmasking of inherited mutations by inherited deletions represents a step toward uncovering missed heritability in MR/MCA,⁵⁴ a recent study has highlighted possible involvement of *de novo* mutations.⁵⁵ Such *de novo* mutations may occur either within or outside a region with a hemizygous deletion and contribute to phenotypes by a recessive or a dominant mechanism, respectively. In cases in which one parent transmits a hemizygous deletion and the other parent carries two homozygous reference alleles, a non-reference allele in the child may be interpreted as a sequencing error, whereas this may reflect a *de novo* mutation on the allele from the other parent. Similarly, transmission of a deletion from one parent and a heterozygous allele from the other parent in the deletion region will appear as homozygous in the child. Without taking into account copy number status, such a situation will also likely be filtered out as a false positive. Therefore, current bioinformatics tools need to be adapted to allow for non-Mendelian constellations^{56,57} or require a combination of experimental techniques that allow for detection of both single-nucleotide variation and copy number status. The results of such studies will also present us with novel challenges during genetic counselling of families.^{58–63}

In conclusion, this study suggests that CNVs may contribute to clinical phenotypes by a recessive mode of gene action. We conclude that array-based enrichment of inherited deletions may be a targeted and highly sensitive method to detect SNVs and CNVs, simultaneously. Our study extends the scope of genome-wide CNV profiling beyond *de novo* CNVs in sporadic patients, and may aid in uncovering missing heritability in genome-wide screening studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the patients and their parents for their generous cooperation. We gratefully acknowledge the technicians of the laboratories of clinical cytogenetics and DNA-diagnostics at the Department of Medical Genetics as well as Marie-Jose van den Boogaard, Eva H Brilstra, Saskia Bulk, Albertien M van Eerde, Jacques C Giltay, Jeske J van Harssel, PF Ippel, Hester Y Kroes, Tom G Letteboer, Klaske D Lichtenbelt, Merel Maiburg, Jasper J van der Smagt and Nienke E Verbeek for contributing patients and phenotypic details. We thank Marcel R Nelen for array CGH data and we are grateful to Rolf Pfundt (Department of Human Genetics, Radboud University, Nijmegen Medical Centre, Nijmegen, The Netherlands) for performing Affymetrix 250K array analysis in case 5 and sharing his results with us.

- idiopathic developmental delay in the Netherlands. *Eur J Med Genet* 2009; **52**: 161–169.
- 2 Shaffer LG, Bejjani BA, Tochla B, Kirkpatrick S, Coppinger J, Ballif BC: The identification of microdeletion syndromes and other chromosome abnormalities: cytogenetic methods of the past, new technologies of the future. *Am J Med Genet Part C* 2007; **145C**: 335–345.
- 3 Xiang B, Xhu H, Shen Y *et al*: Genome-wide oligonucleotide array comparative hybridization for etiological diagnosis of mental retardation. A multicenter experience of 1499 clinical cases. *J Molec Diagn* 2010; **12**: 204–212.
- 4 Hatchwell E: Monozygotic twins with chromosome 22q11 deletion and discordant phenotype. *J Med Genet* 1996; **33**: 261.
- 5 Gardner RJM, Sutherland GR: *Chromosome Abnormalities and Genetic Counseling*. 3rd edn, Oxford University Press: Oxford, 2001.
- 6 Coman DJ, Gardner RJM: Deletions that reveal recessive genes. *Eur J Hum Genet* 2007; **15**: 1103–1104.
- 7 Friedrich K, Lee L, Leistriz DF *et al*: *WRN* mutations in Werner syndrome patients: genomic rearrangements, unusual intronic mutations and ethnic-specific alterations. *Hum Genet* 2010; **128**: 103–111.
- 8 Rujescu D, Ingason A, Cichon S *et al*: Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum Mol Genet* 2009; **18**: 988–996.
- 9 Boone PM, Bacino CA, Shaw CA *et al*: Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 2010; **3**: 1326–1342.
- 10 Quemener S, Chen JM, Chuzhanova N *et al*: Complete ascertainment of intragenic copy number mutations (CNMs) in the *CFTR* gene and its implications for CNM formation at other autosomal loci. *Hum Mutat* 2010; **31**: 421–428.
- 11 Nesteruk M, Peters SU *et al*: Intragenic rearrangements in *NRXN1* in three families with autism spectrum disorder, developmental delay, and speech delay. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153**: 983–993.
- 12 Tsai ACJ, Dosssett C, Walton CSE *et al*: Exon deletions of the *EP300* and *CREBBP* genes in two children with Rubinstein-Taybi syndrome detected by aCGH. *Eur J Hum Genet* 2011; **19**: 43–49.
- 13 Rooryck C, Morice-Picard F, Lasseaux E *et al*: High resolution mapping of *OCA2* intragenic rearrangements and identification of a founder effect associated with a deletion in Polish albino patients. *Hum Genet* 2011; **129**: 199–208.
- 14 Rivera-Brugués N, Albrecht B, Wiczorek D *et al*: Cohen syndrome diagnosis using whole genome arrays. *J Med Genet* 2011; **48**: 136–140.
- 15 Hurst JA, Jenkins D, Vasudevan PC *et al*: Metopic and sagittal synostosis in Greig cephalopolysyndactyly syndrome: five cases with intragenic mutations or complete deletions of *GLI3*. *Eur J Hum Genet* 2011; **19**: 757–762.
- 16 Poot M, Eleveld MJ, van't Slot R, Ploos van Amstel HK, Hochstenbach R: Recurrent copy number changes in mentally retarded children harbour genes involved in cellular localization and the glutamate receptor complex. *Eur J Hum Genet* 2010; **18**: 39–46.
- 17 Poot M, Beyer V, Schwaab I *et al*: Disruption of *CNTNAP2* and additional structural genome changes in a boy with speech delay and autism spectrum disorder. *Neurogenetics* 2010; **11**: 81–89.
- 18 Mokry M, Feitsma H, Nijman IJ *et al*: Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. *Nucleic Acids Res* 2010; **38**: e116.
- 19 Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 2009; **25**: 1754–1760.
- 20 Nijman IJ, Mokry M, van Bostel R, Toonen P, de Bruijn E, Cuppen E *et al*: Mutation discovery by targeted genomic enrichment of multiplexed barcoded samples. *Nat Methods* 2010; **7**: 913–915.
- 21 Grantham R: Amino acid difference formula to help explain protein evolution. *Science* 1974; **185**: 862–864.
- 22 Cooper GM, Stone EA, Asimenos G *et al*: Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 2005; **15**: 901–913.
- 23 Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S: Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* 2010; **6**: e1001025.
- 24 Thomas PD, Campbell MJ, Kejariwal A *et al*: PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* 2003; **13**: 2129–2141.
- 25 Adzhubei IA, Schmidt S, Peshkin L *et al*: A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249.
- 26 Kidd JM, Cooper GM, Donahue WF *et al*: Mapping and sequencing of structural variation from eight human genomes. *Nature* 2008; **453**: 56–64.
- 27 Miller DT, Adam MP, Aradhya *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.
- 28 Barber JCK: Directly transmitted unbalanced chromosome abnormalities and euchromatic variants. *J Med Genet* 2005; **42**: 609–629.
- 29 Girirajan S, Eichler EE: Phenotypic variability and genetic susceptibility to genomic disorders. *Hum Mol Genet* 2010; **19** (R2): R176–R187.
- 30 Sharp AJ: Emerging themes and new challenges in defining the role of structural variation in human disease. *Hum Mutat* 2008; **30**: 135–144.
- 31 Klopocki E., Schulze H., Strauss G *et al*: Complex inheritance pattern resembling autosomal recessive inheritance involving a microdeletion in thrombocytopenia-absent radius syndrome. *Am J Hum Genet* 2007; **80**: 232–240.
- 32 Mefford HC, Sharp AJ, Baker C *et al*: Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *New Eng J Med* 2008; **359**: 1685–1699.
- 33 Hannes FD, Sharp AJ, Mefford HC *et al*: Recurrent reciprocal deletions and duplications of 16p13.11: the deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J Med Genet* 2009; **46**: 223–232.

1 Hochstenbach R, van Binsbergen E, Engelen J *et al*: Array analysis and karyotyping: workflow consequences based on a retrospective study of 36325 patients with

- 34 Kumar R.A., Marshall C.R., Badner J.A *et al*: Association and mutation analyses of 16p11.2 autism candidate genes. *PLoS One* 2009; **4**: e4582.
- 35 Girirajan S, Rosenfeld JA, Cooper GM *et al*: A recurrent 16p12.1 microdeletion supports a two-hit model for severe developmental delay. *Nat Genet* 2010; **42**: 203–209.
- 36 Firth HV, Richards SM, Bevan AP *et al*: DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet* 2009; **84**: 524–533.
- 37 Curry CJ, Mao R, Aston E *et al*: Homozygous deletions of a copy number change detected by array CGH: a new cause for mental retardation? *Am J Med Genet A* 2008; **146A**: 1903–1910.
- 38 Dirks RP, van Geel R, Hensen SM, van Genesen ST, Lubsen NH: Manipulating heat shock factor-1 in *Xenopus* tadpoles: neuronal tissues are refractory to exogenous expression. *PLoS One* 2010; **5**: e10158.
- 39 Pierce A, Wei R, Halade D, Yoo SE, Ran Q, Richardson A: A Novel mouse model of enhanced proteostasis: Full-length human heat shock factor 1 transgenic mice. *Biochem Biophys Res Commun* 2010; **402**: 59–65.
- 40 Khodosevich K, Seeburg PH, Monyer H: Major signaling pathways in migrating neuroblasts. *Front Mol Neurosci* 2009; **2**: 7.
- 41 Newbury DF, Monaco AP: Genetic advances in the study of speech and language disorders. *Neuron* 2010; **68**: 309–320.
- 42 Itsara A, Cooper GM, Baker C *et al*: Population analysis of large copy number variants and hotspots of human genetic disease. *Am J Hum Genet* 2009; **84**: 148–161.
- 43 Cooper GM, Coe BP, Girirajan S *et al*: A copy number variation morbidity map of developmental delay. *Nat Genet* 2011; **43**: 838–846.
- 44 Conrad DF, Pinto D, Redon R *et al*: Origins and functional impact of copy number variation in the human genome. *Nature* 2010; **464**: 704–712.
- 45 Lee ST, Nicholls RD, Bunday S, Laxova R, Musarella M, Spritz RA: Mutations of the *P* gene in oculocutaneous albinism, and Prader-Willi syndrome plus albinism. *New Eng J Med* 1994; **330**: 529–534.
- 46 Fridman C, Hosomi N, Varela MC, Souza AH, Fukai K, Koiffmann CP: Angelman syndrome associated with oculocutaneous albinism due to an intragenic deletion of the *P* gene. *Am J Med Genet Part A* 2003; **119A**: 180–183.
- 47 Liburd N, Ghosh M, Riazuddin S *et al*: Novel mutations of *MYO15A* associated with profound deafness in consanguineous families and moderately severe hearing loss in a patient with Smith-Magenis syndrome. *Hum Genet* 2001; **109**: 535–541.
- 48 Flipsen-ten Berg K, van Hasselt PM, Eleveld MJ *et al*: Unmasking of a hemizygous *WFS1* gene mutation by a chromosome 4p deletion of 8.3 Mb in a patient with Wolf-Hirschhorn syndrome. *Eur J Hum Genet* 2007; **15**: 1132–1138.
- 49 Stec I, Wright TJ, van Ommen GJ *et al*: WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a *Drosophila* dysmorphia gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum Mol Genet* 1998; **7**: 1071–1082.
- 50 Darvish H, Esmaeeli-Nieh S, Monajemi GB *et al*: A clinical and molecular genetic study of 112 Iranian families with primary microcephaly. *J Med Genet* 2010; **47**: 823–828.
- 51 Walsh T, Shahn H, Elkan-Miller T *et al*: Whole exome sequencing and homozygosity mapping identify mutation in the cell polarity protein *GPSM2* as the cause of nonsyndromic hearing loss DFNB82. *Am J Hum Genet* 2010; **87**: 90–94.
- 52 Vermeer S, Hoischen A, Meijer RP *et al*: Targeted next-generation sequencing of a 12.5 Mb homozygous region reveals ANO10 mutations in patients with autosomal-recessive cerebellar ataxia. *Am J Hum Genet* 2010; **87**: 813–819.
- 53 Kuss AW, Garshasbi M, Kahrizi K *et al*: Autosomal recessive mental retardation: homozygosity mapping identifies 27 single linkage intervals, at least 14 novel loci and several mutation hotspots. *Hum Genet* 2011; **129**: 141–148.
- 54 Manolio TA, Collins FS, Cox NJ *et al*: Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
- 55 Vissers LE, de Ligt J, Gilissen C *et al*: A de novo paradigm for mental retardation. *Nat Genet* 2010; **42**: 1109–1112.
- 56 McCarroll SA, Hadnott TN, Perry GH *et al*: Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008; **40**: 1253–1260.
- 57 McCarroll SA, Hadnott TN, Perry GH *et al*: Common deletion polymorphisms in the human genome. *Nat Genet* 2006; **38**: 86–92.
- 58 van Ravenswaaij-Arts CM, Kleefstra T: Emerging microdeletion and microduplication syndromes; the counseling paradigm. *Eur J Med Genet* 2009; **52**: 75–76.
- 59 Koolen DA, Pfundt R, de Leeuw N *et al*: Genomic microarrays in mental retardation: a practical workflow for diagnostic applications. *Hum Mutat* 2009; **30**: 283–292.
- 60 Gijsbers AC, Lew JY, Bosch CA *et al*: A new diagnostic workflow for patients with mental retardation and/or multiple congenital abnormalities: test arrays first. *Eur J Hum Genet* 2009; **17**: 1394–1402.
- 61 Buysse K, Delle Chiaie B, Van Coster R *et al*: Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. *Eur J Med Genet* 2009; **52**: 398–403.
- 62 Poot M, Hochstenbach R: A three-step workflow procedure for the interpretation of array-based comparative genomic hybridization results in patients with idiopathic mental retardation and congenital anomalies. *Genet Med* 2010; **12**: 478–485.
- 63 Poot M, van der Smagt JJ, Brilstra EH, Bourgeron T: Disentangling the myriad genomics of complex disorders specifically focusing on autism, epilepsy and schizophrenia. *Cytogenet Genome Res* 2011; **135**: 228–240.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)