

ARTICLE

Single exon-resolution targeted chromosomal microarray analysis of known and candidate intellectual disability genes

Tracy Tucker^{1,7}, Farah R Zahir^{*,2,3,7}, Malachi Griffith³, Allen Delaney³, David Chai¹, Erica Tsang², Emmanuelle Lemyre⁴, Sylvia Dobrzaniecka⁴, Marco Marra³, Patrice Eydoux¹, Sylvie Langlois^{2,5}, Fadi F Hamdan⁴, Jacques L Michaud⁴ and Jan M Friedman^{2,6}

Intellectual disability affects about 3% of individuals globally, with ~50% idiopathic. We designed an exonic-resolution array targeting all known submicroscopic chromosomal intellectual disability syndrome loci, causative genes for intellectual disability, and potential candidate genes, all genes encoding glutamate receptors and epigenetic regulators. Using this platform, we performed chromosomal microarray analysis on 165 intellectual disability trios (affected child and both normal parents). We identified and independently validated 36 *de novo* copy-number changes in 32 trios. In all, 67% of the validated events were intragenic, involving only exon 1 (which includes the promoter sequence according to our design), exon 1 and adjacent exons, or one or more exons excluding exon 1. Seventeen of the 36 copy-number variants involve genes known to cause intellectual disability. Eleven of these, including seven intragenic variants, are clearly pathogenic (involving *STXBP1*, *SHANK3* (3 patients), *IL1RAPL1*, *UBE2A*, *NRXN1*, *MEF2C*, *CHD7*, 15q24 and 9p24 microdeletion), two are likely pathogenic (*PI4KA*, *DCX*), two are unlikely to be pathogenic (*GRIK2*, *FREM2*), and two are unclear (*ARID1B*, 15q22 microdeletion). Twelve individuals with genomic imbalances identified by our array were tested with a clinical microarray, and six had a normal result. We identified *de novo* copy-number variants within genes not previously implicated in intellectual disability and uncovered pathogenic variation of known intellectual disability genes below the detection limit of standard clinical diagnostic chromosomal microarray analysis.

European Journal of Human Genetics (2014) 22, 792–800; doi:10.1038/ejhg.2013.248; published online 20 November 2013

Keywords: targeted; clinical; CMA; intragenic; pathogenic

INTRODUCTION

Intellectual disability (ID) affects about 3% of individuals globally, and, for half the cases, the cause is unknown. The wide-spread use of chromosomal microarray analysis (CMA) led to a new frontier in clinical diagnosis, with the ability to detect causative submicroscopic chromosomal imbalances, also called pathogenic copy-number variants (CNVs), in at least 10–15% of affected individuals in whom conventional cytogenetic analysis is normal.^{1–3}

A few years ago, the number of genes recognized to contribute to ID was reported as ~300.^{4,5} Currently, the estimated number of ID genes is at least 900–950, based on the evidence that there are 91 pathogenic X-linked genes that account for 10–15% of ID in males.⁶

Although the power of next-generation sequencing has opened up the potential to screen all exons or the whole genome for sequence mutations, the analysis algorithms are not, as yet, robust for routine identification of losses or gains of sequence involving a single or a few exons. As the probe capacity on microarrays increases, there is the potential to screen a large number of genes at the exonic level using microarrays^{7–10} providing the means to identify CNVs that are

currently missed with whole-genome clinical arrays and next-generation sequencing.

We designed a microarray with a single-exon resolution and screened 1397 genes known or hypothesized to cause ID. We screened 165 trios composed of a child with idiopathic ID and both normal parents using this array. We detected and independently validated 36 CNVs in 32 families. Seventeen of these involve genes known to cause ID, of which at least 11, including 7 that are intragenic, are clearly pathogenic. Our results confirm the efficacy of our design and offer novel insights into the pathogenesis of ID.

MATERIALS AND METHODS

Subjects

Patients with ID with or without additional clinical features were selected for study. The cause of the ID in each child was unknown despite full evaluation by a clinical geneticist, a karyotype at ≥ 500 band resolution and subtelomeric FISH studies. Autism was diagnosed using the Autism Diagnostic Schedule/Autism Diagnostic Interview (ADOS/ADI). This study was approved by the University of British Columbia Clinical Research Ethics Board and Sainte-Justine Hospital Ethics Board. Informed consent was obtained for each patient.

¹Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada; ²Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; ³Canada's Michael Smith Genome Sciences Centre, Vancouver, British Columbia, Canada; ⁴CHU Sainte-Justine Research Center, Montréal, Quebec, Canada; ⁵Provincial Medical Genetics Programme, Children's & Women's Health Centre of British Columbia, Vancouver, British Columbia, Canada; ⁶Child & Family Research Institute, Vancouver, British Columbia, Canada

⁷Joint first authors.

*Correspondence: Dr FR Zahir, Department of Medical Genetics, University of British Columbia, Children's and Women's Hospital, Box 153, 4500 Oak Street, Vancouver, British Columbia, Canada V6H 3N1. Tel: +604 875 2000 ext. 5625; Fax: +604 875 3019; E-mail: farahz@cfri.ca

Received 20 March 2013; revised 5 September 2013; accepted 27 September 2013; published online 20 November 2013

Paternity and maternity were confirmed by using six highly informative unlinked microsatellite markers, as previously described.¹¹

Custom array design

Our goal was to design a custom NimbleGen 12-plex array with 135 000 probes covering suspected or known genes involved in the development of ID at time of design (April 2008), with a minimum of eight probes per exon in each of our selected genes. The NCBI human genome sequence build 36.1 was used as the reference sequence. Detailed array design is provided in Supplementary Notes S1, and the complete array and data discussed have been deposited into the NCBI Gene Expression Omnibus (GEO) repository and are publicly available (<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE39533>). Briefly, the study comprised six main stages: (a) selection of genes and design of oligonucleotide probes, (b) testing of a pilot phase NimbleGen 385 K array with validation on control samples from patients with known CNVs, (c) selection of best performing probes and design of the final NimbleGen 135 K array, (d) CMA of 165 idiopathic ID trios (affected child and both unaffected parents) using the final NimbleGen 12-plex array (each subarray containing 135 K probes) and bioinformatic analysis to identify *de novo* CNVs, (e) validation of CMA results by quantitative PCR using SYBR green, and (f) genotype–phenotype correlation to determine pathogenic relationship of validated *de novo* CNVs to ID. The following genes were selected for inclusion on the array: (1) genes previously shown to cause ID (obtained from OMIM with keywords ‘mental retardation’ or ‘intellectual disability’; October 2007), (2) all genes within reported microdeletion/microduplications (<100 kb), reported in the DECIPHER database (October 2007), in which the phenotype included ID, (3) ID candidate genes within reported microdeletion/microduplications (>100 kb) reported in the DECIPHER database (October 2007), in which the phenotype included ID. If there were no reported candidate genes within these CNVs, a set of eight probes was placed every 10 kb within genes or highly conserved regions, (4) all brain-expressed glutamate receptors (GRC) and the majority of their known interacting proteins,¹² and (5) genes involved in epigenetic regulation. A complete list of genes and regions covered by the array can be found in Supplementary Notes (ST1). The selection resulted in a total of 1397 RefSeq genes. The probes were selected according to a previously published protocol,¹³ and the arrays were synthesized by NimbleGen according to our custom design specification. Figure 1 shows the location of all probes and coverage obtained by our final 135 K array design.

Array hybridization and data analysis

The labeling, hybridization and washing of the array were performed according to the manufacturer’s specifications, and analysis was done with default settings using the NimbleScan software version 2.1.4. (Supplementary Notes S1). CNVs were identified when the \log_2 ratio was >0.2 (duplications) or <−0.2 (deletions) with a minimum of five probes affected, using the BioDiscovery Nexus Software. The minimally affected region within a CNV was considered to be between the location of the first and last probe with an abnormal \log_2 ratio (>0.2 or <−0.2). The maximally affected region for a CNV was defined as the region between the location of the normal probe (\log_2 ratio between ± 0.2) immediately preceding the most proximal abnormal probe and the normal probe immediately following the most distal abnormal probe.

Defining *de novo* CNVs

The analyses measured the strength of the hybridization signal obtained in two different array genomic hybridization experiments – one with the child’s DNA in comparison with that of the mother and the second with the child’s DNA in comparison with that of the father. A CNV was considered to be *de novo* if it was independently identified in both hybridizations. All inherited variants (CNVs that were seen on a hybridization versus one parent but not the other) were analyzed visually, and the CNV was reclassified as *de novo* if there appeared, by eye, to be a shift of the probe \log_2 ratio in the same direction in the second parent that had not been called by the software. Each CNV call in which the child’s signal was less than that of the parent (called a ‘loss’ in the tally) could actually represent either a loss of copy-number in the child or a gain of copy-number in the parent. Similarly, each CNV call in which the child’s signal was greater than that of the parent (called a ‘gain’ in the tally)

could actually represent either a gain of copy-number in the child or a loss of copy-number in the parent. The direction of the CNV was confirmed with independent qPCR validation on the trio using a commercially available pooled reference set (see Supplementary Notes S1). We have used CNV as a general term without a size limitation because of current recommendations that the original 1 kb minimum CNV size was a reflection of early technologies.^{14,15} Exons were numbered according to the NCBI reference sequence database (RefSeq).

Validation of *de novo* CNVs

CNVs were validated by qPCR ($\Delta\Delta C_t$ method) using SYBR Green (Applied Biosystems, Life Technologies, Carlsbad, CA, USA) on an ABI 7500 fast real-time PCR system using both parents and a pooled reference sample from Promega (Madison, WI, USA) (Catalog#: G3041 – male and female and G1521-female only) with hexose-6-phosphate dehydrogenase (*H6PD* MIM#138090) as the control locus. Primers (listed in ST2) were designed using Primer Express (Applied Biosystems) (detailed in S1). CNVs on the X chromosome were validated against a pooled reference sample composed of female only DNA, whereas those on autosomes were validated against a pooled reference sample composed of both male and female DNA.

Pathogenicity

A *de novo* CNV was classified as pathogenic when all of the following criteria were met: (1) it occurred within a known ID gene, (2) was predicted to disrupt the gene, and (3) the reported phenotype and the phenotype in our patient overlapped. A *de novo* CNV was classified as likely pathogenic when our patients’ phenotype and the phenotype reported with disruptions in the gene overlapped, but the CNV we identified has not been previously reported.

RESULTS

We designed a custom array targeting known and hypothesized ID genes with sufficient probe coverage within exons to detect a CNV involving a single exon (Figure 1). We used this custom platform to perform CMA in 165 trios, each consisting of a child with idiopathic ID and both unaffected parents. We identified 176 putative *de novo* CNVs, of which 36 in 32 trios were confirmed by qPCR (Tables 1 and 2). The ~21% validation rate we achieved is in keeping with other CMA research studies using whole-genome arrays.¹⁶ Four individuals (patients 931, 331, 600, and 513) exhibited two *de novo* CNVs each. For each of these four patients, one CNV was clearly pathogenic and the contribution, if any, of the second CNV is unknown.

Of the 36 confirmed *de novo* CNVs (27 deletions and 9 duplications), nine involved at least the whole gene (with the possible maximally affected region also including adjacent genes) and three involved large regions known or strongly suspected to cause syndromic ID (15q24 microdeletion, 15q22 microdeletion, and a 5-Mb deletion of 9p24). We confirmed 15 (42%) *de novo* intragenic single or multi-exon CNVs (examples given in Figure 2), of which seven CNVs (five deletions and two duplications) in seven patients occurred in the *JARID2* gene – further studies suggest this is a benign polymorphism occurring at high population frequency (manuscript in preparation). Seven additional CNVs (19%) involve either only exon 1 or exon 1 and the adjacent few exons, and all were analyzed for involvement of the translational start site or promoter region if probe coverage was present upstream of exon 1 using FirstEF, an *in silico* program.¹⁷ Of these seven, three CNVs in two genes (*ARID1B*, *CDH2*) removed the translational start site, one CNV involving exon 1 (*CHD6*) removed the predicted promoter sequence and another CNV (*SNTG2*) involved only the upstream regulatory sequence of a gene including the promoter as predicted by FirstEF.

We identified 14 CNVs in 14 individuals involving 12 genes, known to be causative for ID, as well as three microdeletion CNVs (in three individuals). Table 1 summarizes our array findings and the clinical

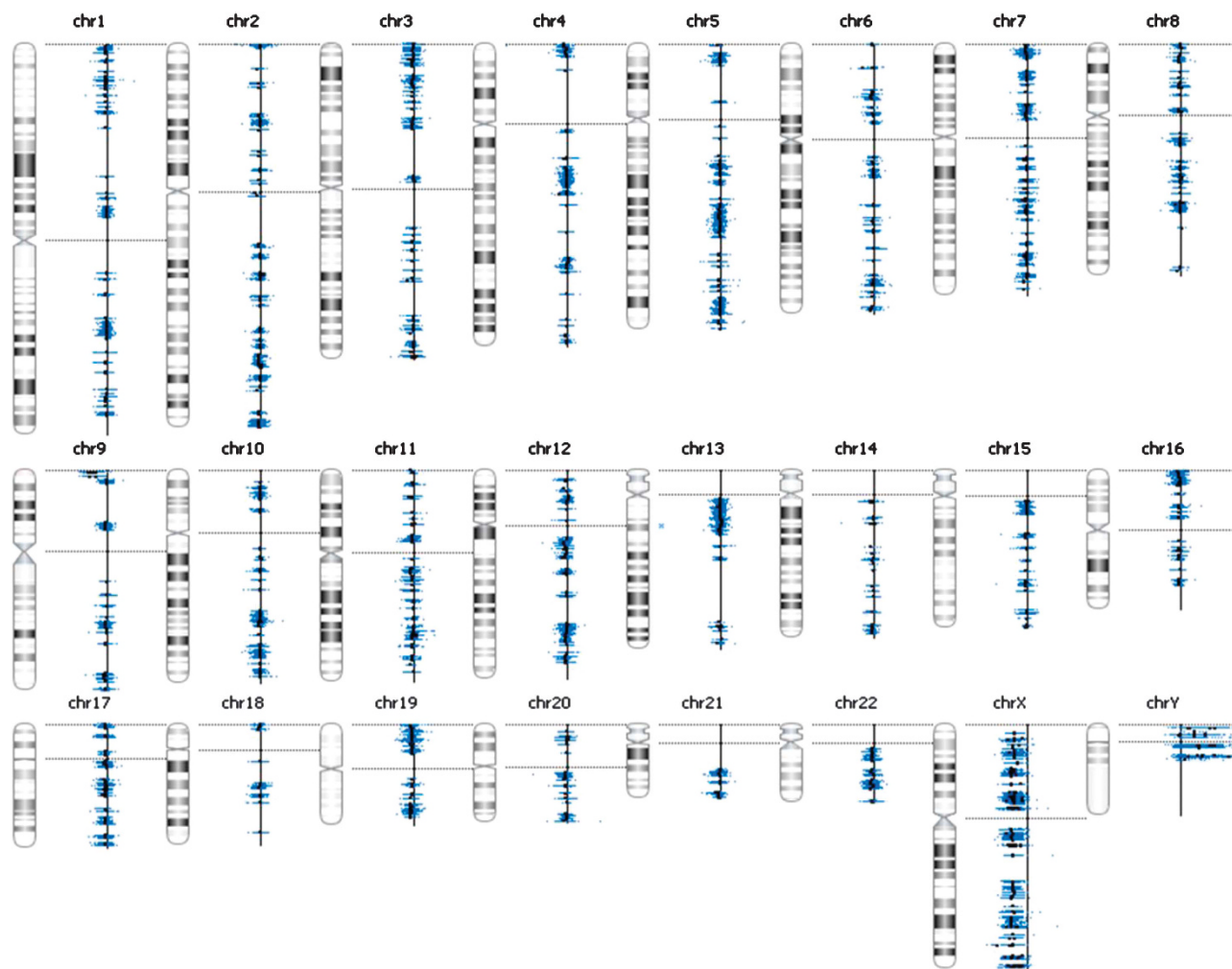


Figure 1 Location of probes against a chromosome ideogram. The blue dots clustered around the black vertical line represent individual probes, with the vertical line representing the normal \log_2 ratio signal of 0. In this case the patient has a duplication of chromosome 9pter as seen by a shift to the left from the normal by the vertical cluster of probes at that region. In addition, the hybridization in this case is of a female child versus the father, as can be seen by a shift toward the left, signifying a gain, from the normal of all chromosome X probes and shift to the right from the normal, signifying a loss, of all chromosome Y probes.

phenotype of these patients. We consider 11 CNVs to be pathogenic in our patients based on the loss or predicted loss of function of the encoded protein and the phenotypic overlap between our patients and those previously reported (*STXBPI*,^{11,18} *SHANK3* (three patients),¹² *IL1RAPL1*,^{19,20} *UBE2A*,^{21,22} *NRXN1*,^{23,24} *MEF2C*,²⁵ *CHD7*,^{26,27} 15q24²⁸ and 9p24 microdeletions). We consider two whole gene duplications to be likely pathogenic (patients 224, 8327); *DCX* duplication has not been reported in ID, but is sensitive to loss,²⁹ the *PI4KA* duplication in our patient has already been published.³⁰ We identified two CNVs affecting genes (*FREM2*, *GRIK2*) associated with autosomal recessive forms of ID. In patient 419 with a heterozygous deletion of *GRIK2*, Sanger sequencing of all *GRIK2* exons and their intronic boundaries did not reveal any mutations making it unlikely that the phenotype in this patient was due to loss of *GRIK2*. The phenotype of patient 354 (Table 1) was not consistent with that reported for autosomal loss of *FREM2*. Therefore, we considered these two CNVs unlikely to be pathogenic in these patients. For two other *de novo* CNVs, pathogenicity is unclear (*ARID1B*, 15q22 microdeletion). We identified a duplication of exon 1 of *ARID1B*, and the genomic location of this extra exon is

unknown. The effect of a tandem duplication of exon 1 is difficult to predict, as in this case it does not disrupt the reading frame and would require functional studies to determine pathogenicity. In the case of the 15q22 deletion, the minimally effected region contains three non-ID genes (*GCNT3*, *FOXB1* and *BNIP2*) and maximally 13.8 Mb of sequence maybe involved, therefore, without a whole-genome CMA to better delineate the extent and genes involved in the CNV, pathogenicity is unclear.

CMA using whole-genome research or clinical arrays were run on nine of the patients who had confirmed *de novo* CNVs within genes known to cause ID (Table 1). Of these nine cases, six were also abnormal by clinical array with confirmed CNVs that overlapped with our affected regions, while three of the *de novo* CNVs we found were not identified by clinical CMA. The first of these was an 18.8-kb pathogenic deletion of exons 20–23 of *SHANK3* (patient 828). Interestingly, the same clinical array identified a 43-kb deletion of exons 9–23 of *SHANK3* in two other patients (patients 622 and 248), in whom our array also detected the same CNVs. The second was a 3.5-kb deletion of exons 10 and 11 of *STXBPI* (patient 970) and the third was a 5.5-Mb deletion involving the whole *NRXN1*

Table 1 Summary of confirmed *de novo* CNVs identified in genes or microdeletion/microduplication regions previously associated with ID and summary of patient phenotype and other CMA information

Patient ID	CNV	Minimum size (Maximum Size) Kb	Gene			Patient phenotype	DGV entries mapping to coding region of named gene(s)	Clinical or research CMA result (pathogenicity)
			within min region (MIM #)	Minimum gene region affected	Expected effect of CNV on protein ^a			
419	chr6.hg18:g.(102,483,616_102,589,910)_ (102,590,200_102,609,493)del	0.3 (125.9)	<i>GRIK2</i> (611092)	Exon 14	Frameshift/ premature STOP	Non-syndromic ID	None	None (Unlikely pathogenic)
331 ^b	chr6.hg18:g.(157,133,828_157,141,000)_ (157,142,500_157,192,105)dup	1.5 (58.3)	<i>ARID1B</i> (614556)	Exon 1	Includes translational start site	ID, hypotonia, gross motor delay, cataract, dysmorphic facies	One DGV entry (#111727) showing gain within exon 1	None (Unclear pathogenicity)
331 ^b	chr8.hg18:g.(61,748,366_61,753,950)_ (61,754,150_61,816,132)	0.2 (67.8)	<i>CHD7</i> (608692)	Exon 1	Effect of 6 copy gain unpredictable	ID, cataract	One small variant mapping to intron of <i>CHD7</i>	None (Pathogenic)
970	chr9.hg18:g.(129,470,063_129,470,133)_ (129,473,690_129,473,728)del	3.6 (3.7)	<i>STXBP1</i> (612164)	Exon 10–11	Frameshift/ premature STOP	ID, infantile spasms		Normal by 135 K Signature Genomics (Pathogenic)
622	chr22.hg18:g.(49,468,843_49,469,360)_ (49,501,100_49,568,694)del	31.7 (99.9)	<i>SHANK3</i> (606232)	Exon 9–23	Removes major functional elements.	ID, severe language impairment	DGV #5193, 4141, 239 completely overlap gene. 6 other small DGV variants within introns of gene	43 Kb deletion by 135 K Signature Genomics (Pathogenic)
248	chr22.hg18:g.(49,468,792_49,469,360)_ (49,501,100_49,568,694)del	31.8 (99.9)	<i>SHANK3</i> (606232)	Exon 9–23	Removes major functional elements.	ID		43 Kb deletion by 135 K Signature Genomics (Pathogenic)
828	chr22.hg18:g.(49,497,450_49,499,800)_ (49,518,600_49,552,931)del	18.8 (55.5)	<i>SHANK3</i> (606232)	Exon 20–23	Removes major functional elements.	ID		Normal by 135 K Signature Genomics (Pathogenic)
881	chrX.hg18:g.(29,038,856_29,178,350)_ (29,211,600_29,323,944)del	33.3 (285.1)	<i>IL1RAPL1</i> (300143)	Exon 3	Frameshift/ premature STOP	ID	DGV #3265 a 3.6 Mb gain	102 Kb deletion by 135 K Signature Genomics (Pathogenic)
600 ^b	chrX.hg18:g.(118,593,066_118,598,210)_ (118,601,900_119,448,574)del	3.7 (855.5)	<i>UBE2A</i> (312180)	Exon 3–6	Frameshift/ premature STOP	ID	None	Normal (Pathogenic)
513 ^b	chr2.hg18:g.(49,069,733_49,999,147)_ (51,006,648_54,601,328)del	1007.5 (5531.6)	<i>NRXN1</i> (600565)	Whole gene	Deleted	ID, autism	None	Normal by 135 K Signature Genomics (Pathogenic)
931 ^b	chr5.hg18:g.(87,100,982_87,412,857)_ (88,163,700_88,257,129)del	750.8 (1156.2)	<i>MEF2C</i> (600662)	Whole gene	Deleted	ID, spasticity, epilepsy, acquired microcephaly	10 small variants, all located within intronic sequence, except for variant # 93139, which involves the final exon.	1 Mb deletion by 135 K Signature Genomics (Pathogenic)
354	chr13.hg18:g.(36,521,031_36,781,657)_ (38,348,500_38,509,610)del	1566.8 (1988.6)	<i>FREM2</i> (219000)	Whole gene	Deleted	ID, congenital microcephaly, cerebellar atrophy	Two small variants mapping to intron	1.5 Mb deletion by 44 K Gene Dx array (Unlikely pathogenic)
8327	chr22.hg18:g.(18,151,442_19,502,640)_ (19,706,747_20,378,348)dup	204.1 (2226.9)	<i>PI4KA</i> (600286)	Whole gene	Duplicated	ID, small sta- ture, Pierre Robin sequence with cleft palate	DGV #73751, 73748, 73747, 79458 map to gene including coding region. Many DGV entries scattered throughout maximally affected region	378 Kb deletion by Agilent, Affymetrix and Nimblegen (Likely pathogenic)

Table 1 (Continued)

Patient ID	CNV	Minimum size (Maximum Size) Kb	Gene			Expected effect of CNV on protein ^a	Patient phenotype	DGV entries mapping to coding region of named gene(s)	Clinical or research CMA result (pathogenicity)
			within min region (MIM #)	Minimum gene region affected	Expected				
224	chrX.hg18:g.(108,863,590_110,225,907)_ (110,541,300_115,215,508)dup	315.4 (6354.9)	DCX (300121)	Whole gene	Duplicated	ID	None	1.7 Mb deletion by Affymetrix 6.0 (Likely pathogenic)	
747	chr15.hg18:g.(72,497,977_72,498,044)_ (73,841,453_77,040,887)del	1343.4 (4542.9)	Del 15q24	1.3 Mb	Large deletion	ID, growth retardation, microcephaly, duodenal atresia Duane abnormality, craniofacial abnormalities	Many DGV entries.	None (Pathogenic)	
452	chr15.hg18:g.(50,608,939_57,695,626)_ (58,155,227_64,466,174)del	459.6 (13857.2)	Del 15q22	460 Kb	Large deletion	ID	Many DGV entries.	None (Unclear pathogenicity)	
412	chr9.hg18:g.(?_2,005,341)_ (2,183,623_6,649,166)del	178 Kb (6649.2)	Del 9p24	178 Kb	Large deletion	ID	Several DGV entries showing duplications (#53551, 95804) and deletions	5 Mb deletion by 135 K Signature Genomics (Pathogenic)	

Abbreviations: CMA, chromosomal microarray, CNV, copy-number variant; DGV, database of genomic variants. CNVs are reported as per Human Genome Variation Society guidelines. Minimal gene region affected refers to the gene content contained within the minimally deleted/duplication region. The minimum size is defined as the difference between the genomic coordinates of the last abnormal probe and the first abnormal probe, and the maximum size is defined as the difference between the genomic coordinates of the first normal probe location distal to the CNV and the last normal probe location proximal to the CNV. Exon numbering is according to the NCBI reference sequence database (RefSeq). http://dgvbeta.tcag.ca/gb2/gbrowse/dgv2_hg18/.

^aEffect on protein predicted using Alamut V2.2.

^bPatients with more than one CNV identified in this study.

gene (patient 513). We were perplexed by the clinical CMA missing this large CNV, despite the platform used probing this gene with 30 markers. It is possible the CNV only includes *NRXN1* and that 30 markers is insufficient to call a CNV, and these findings highlight how CMA data analyses methods may differently affect results.

With regards to CNVs we identified in known pathogenic regions; three patients were found to have CNVs in known pathogenic microdeletion regions. The first was a 1.3- to 4.5-Mb deletion (minimally-maximally affected size) within 15q24, a known microdeletion syndrome region.²⁸ The phenotype of our patient (Table 1) is consistent with the reported syndrome, indicating that this CNV is pathogenic. The second patient (patient 412) had a 178-kb to 6.6-Mb deletion of 9p24 that includes the *SMARCA2* gene targeted by our array. A clinical CMA (with a genomic backbone and thus able to better define breakpoints than our design) characterized this *de novo* CNV as a 5-Mb deletion. The clinical CMA in this patient identified an additional 3.5-Mb duplication CNV of 16q24.1 that we didn't detect due to a lack of probe coverage in the region as per our array design. The clinical report for this patient called the 9p24 deletion pathogenic based on size, while the smaller 16q24.1 CNV was of unknown clinical significance. In the third patient (452), we identified a 450-kb to 13.8-Mb deletion of 15q22, and for reasons explained above we are currently uncertain of the pathogenicity of this event.

In addition to designing a proof-of-principle exon-resolution targeted ID array, one goal of this project was to investigate genes involved in epigenetic regulation and synaptogenesis for involvement in causative CNVs in patients with ID. We identified and independently validated 19 *de novo* CNVs that included genes of these categories in 19 patients (Table 2). For these cases, many of the CNVs that we identified occurred within genes in which single case reports

have been published or show mouse model data consistent with a role in ID. In these instances, further study is necessitated to determine the extent of these CNVs with detailed phenotype-genotype correlations (manuscripts in preparation). We have therefore omitted a detailed assessment of the pathogenicity of these novel findings in this paper.

Interestingly, three of the above 19 patients had a clinical CMA performed, and all of them were normal on the clinical platform. Patient 164, who had a deletion involving at least the *CADM2* gene, was analyzed on a 105K Gene Dx array. We do not know the probe coverage of this gene by the Gene Dx platform; however, it is unlikely to have contained sufficient probes to detect a single gene imbalance as *CADM2*, as it is not part of a recognized ID syndrome. Two other patients (patients 576 and 7581) both showed a minimal loss of exons 1–2 of *CDH2*. One of these patients was part of another CMA study,³⁰ by which no corresponding CNV was identified using an Affymetrix 500 K, Agilent 244 K or NimbleGen 385 K array. However, none of these platforms included a sufficient number of probes (> 5) to call a CNV in this gene. This suggests that the affected region is probably restricted to the deletion of exon 1–2 of *CDH2*, which does remove the translational start site located in exon 1. The third clinical array was run on a patient with a deletion of exon 1 of the *CHD6* gene; however, this was a clinical BAC array (Signature Genomics- the array did not have *CHD6* in its target list as of April 2010). These data highlight the potential for gene centric array design to identify CNVs that could be missed by clinical CMA.

DISCUSSION

The use of whole-genome CMA to identify microdeletion/microduplications has greatly improved the rate of diagnosis of genetic imbalance in ID patients, so much so, hence CMA is now

Table 2 Summary of confirmed *de novo* CNVs identified in genes/loci not previously associated with ID and summary of patient phenotype and other CMA information

Patient ID	CNV	Minimum size (Maximum size) Kb	Gene		Expected effect of CNV on protein ^a	Patient phenotype	DGV entries mapping to coding region of named gene(s)	Clinical or research CMA result
			within Min Region (MIM #)	Minimum gene region affected				
413	chr1.hg18:g.(233,491,105_233,556,700)_(233,557,700_233,560,656)dup	1.0 (69.6)	<i>ARID4B</i> (609696)	Exons 1,2	No frame-shift with tandem duplication	ID, autism	DGV #48278 gain variant, begins upstream of gene and includes up to exon 7 of gene.	
543	chr2.hg18:g.(86,535,860_86,537,100)_(86,537,200_86,537,443)dup	0.3 (74.1)	<i>KDM3A</i> (611512)	Exon 6	Frameshift/premature STOP	ID	None.	Normal (105 K Gene DX)
447	chr2.hg18:g.(209,997,540_210,080,600)_(210,306,600_213,110,930)del	226 (3113)	<i>MAP2</i> (157130)	Whole gene	Whole-gene deletion	ID, autism	One small variant mapping to intron of MAP2	
839	chr2.hg18:g.(853,846_895,990)_(896,180_935,391)del	0.02 (81.5)	<i>SNTG2</i> (608715)	1 kb upstream	Involves predicted promoter	N/A	Many	
164	chr3.hg18:g.(53,954,275_85,100,000)_(86,000,000_121,027,757)del	0.9 (67073.5)	<i>CADM2</i> (609938)	Whole gene	Whole-gene deletion	Global developmental delay	DGV #53089 is an 88 Kb loss in 2 individuals	
513 ^b	chr4.hg18:g.(5,874,387_5,908,550)_(5,908,900_6,114,729)del	0.35 (240.3)	<i>CRMP1</i> (602462)	Whole gene	Whole-gene deletion	ID, autism	11 small DGV entries in maximally affected region, 1 affects coding sequence, rest are intronic/non-genic	
391	chr6.hg18:g.(15,568,665_15,576,500)_(15,577,250_15,595,501)del	0.8 (26.8)	<i>JARID2</i> (601594)	Exon 6	Deletion/tandem duplication does not cause a frame-shift. Exon 6 codes a low complexity motif without known function	ID, short stature, ASD	None	
9390	chr6.hg18:g.(15,568,665_15,576,500)_(15,577,250_15,595,501)del	0.8 (26.8)	<i>JARID2</i>	Exon 6		ID, CL/P, deafness, facial dysmorphisms		
901	chr6.hg18:g.(15,568,665_15,576,500)_(15,577,250_15,595,501)del	0.8 (26.8)	<i>JARID2</i>	Exon 6		ID, overgrowth		
931 ^b	chr6.hg18:g.(15,568,665_15,576,500)_(15,577,250_15,595,501)del	0.8 (26.8)	<i>JARID2</i>	Exon 6		ID, spasticity, epilepsy, acquired microcephaly		
268	chr6.hg18:g.(15,568,665_15,576,500)_(15,577,250_15,595,501)del	0.8 (26.8)	<i>JARID2</i>	Exon 6		ID, autism		
490	chr6.hg18:g.(15,568,665_15,576,500)_(15,577,250_15,595,501)dup	0.8 (26.8)	<i>JARID2</i>	Exon 6		ID with ASD		
341	chr6.hg18:g.(15,568,665_15,576,500)_(15,577,250_15,595,501)dup	0.8 (26.8)	<i>JARID2</i>	Exon 6		ID with minor facial dysmorphisms		
141	chr8.hg18:g.(22,048,373_24,826,912)_(24,832,700_26,491,057)del	5.8 (4442.7)	<i>NEFM</i> (162250)	Whole gene	Whole-gene deletion	ID	None in affected exons but several in maximally affected region	
600 ^b	chr10.hg18:g.(64,647,390_64,698,700)_(64,699,000_64,721,535)dup	0.3 (74.1)	<i>JMJD1C</i> (604503)	Exon 4	Discrepancy of exon size from source curation. Manual curation: no frame-shift. Alamut curation: Frameshift/premature STOP	Non-syndromic ID	None	

Table 2 (Continued)

Patient ID	CNV	Minimum size (Maximum size) Kb	Gene within		Expected effect of CNV on protein ^a	Patient phenotype	DGV entries mapping to coding region of named gene(s)	Clinical or research CMA result
			Minimum Region (MIM #)	Minimum gene region affected				
7484	chr12:hg18:g.(44,409,886_44,409,000)_ (44,412,000_44,588,086)dup	3.0 (178.2)	<i>ARID2</i> (609539)	Exons 1,2,3 (6 copy gain by qPCR)	Effect of six copy gain unpredictable. Single-tandem duplication causes frame-shift and premature STOP	ID, facial dysmorphism, short stature, unilateral deafness	DGV # 3887 gain variant begins before exon 3 and extends beyond gene.	
576	chr18:hg18:g.(23,982,030_24,010,750)_ (24,011,500_37,690,792)del	0.8 (13708.8)	<i>CDH2</i> (114020)	Exon 1,2	Removes transcrip- tional start site	Global devel- opmental delay	None	
7581	chr18:hg18:g.(23,982,030_24,010,750)_ (24,011,500_37,690,792)del	0.8 (13708.8)	<i>CDH2</i> (114020)	Exon 1,2	Removes transcrip- tional start site	ID, cleft lip/ palate	See above	Normal (Affymetrix 500 K, Agilent 244 K, NimbleGen 385 K)
706	chr20:hg18:g.(39,613,550_39,680,100)_ (39,680,750_39,688,521)del	0.7 (75.0)	<i>CHD6</i>	Exon 1		ID, autism, epilepsy	Two small variants mapped to introns.	Normal (BAC array Signature Genomics)

Abbreviations: CMA, chromosomal microarray; CNV, copy-number variant; DGV, database of genomic variants. CNVs are reported as per Human Genome Variation Society guidelines. Minimal gene region affected refers to the gene content contained within the minimally deleted/duplication region. Minimal size is defined as the difference between the genomic coordinates of the last abnormal probe and the first abnormal probe, and the maximum size is defined as the difference between the genomic coordinates of the first normal probe location distal to the CNV and the last normal probe location proximal to the CNV. Exon numbering is according to the NCBI reference sequence database (RefSeq). http://dgvbeta.tcag.ca/gb2/gbrowse/dgv2_hg18/.

^aEffect on protein predicted using Alamut V2.2.

^bPatients with more than one CNV identified in this study.

recommended as the first line test for individuals with ID.³¹ There has been an increasing interest in intragenic CNVs, and a number of investigators have reported single-exon-resolution CMA for small sets of carefully chosen genes in affected individuals^{7–10} and in normal individuals.^{32,33} However, there have been few studies assessing a large set of candidate genes thought to be causative for a complex and highly heterogeneous condition like ID.³⁴ Therefore, the aim of our study was to design a custom array able to identify CNVs in known ID genes/loci as well as in candidate ID genes, at single-exon resolution, which is well below the current level of detection of standard clinical CMA.

Of the 36 validated *de novo* CNVs, 23 were intragenic, involving one or more exons within a single gene (Tables 1 and 2). Boone *et al.* reported results from their analysis of 3743 cases referred to the Medical Genetics Laboratory at Baylor College of Medicine for CMA with a custom targeted clinical array probing ~1700 candidate genes for a variety of clinical conditions including ID, at an average coverage of four probes per exon.³⁵ The authors found 40 CNVs involving one or more exons of which 15 were known to cause a recognized phenotype concordant with that of the patients. Three of the genes involved in *de novo* CNVs in our study – *IL1RAP1*, *STXBPI*, and *NRXN1* – were also identified in the Baylor cohort, although no overlap in intragenic deletions was observed. Similar to the study by Boone *et al.*, our study highlights the high proportion, 64% in our study, of intragenic CNVs that are pathogenic or potentially pathogenic for disease.

Whole-genome CMA identifies pathogenic CNVs in 15–20% of children with ID who have a normal karyotype.⁶ We hypothesized that we could identify more pathogenic CNVs with a targeted exon resolution array than with a whole-genome or clinical CMA.

As reported above, 12 of our patients with validated *de novo* CNVs were also tested on various other clinical or whole-genome research CMA platforms; for six patients, the other platforms reported normal results. The CNVs in three of these patients we consider to be pathogenic, and all are intragenic CNVs involving genes known to cause ID. Of the six known ID loci that were also detected by clinical CMA, the average size of the CNV was 2600 kb (range 43 kb–5000 kb), which was significantly larger than the average size (9.3 kb; range 3.5 kb–18.5 kb) of the three CNVs involving known ID genes that were missed by the clinical CMA. Current whole-genome clinical arrays have sparse coverage of individual exons within a gene and rely on probe coverage within the whole gene to identify a CNV. This type of design would miss single exon or possibly multi-exon CNVs. In our array design, we chose to have eight probes per exon and used a minimum of five probes showing a shift of the log₂ ratio to be called as a CNV enabling us to detect intragenic CNVs that are missed by platforms with less dense probe coverage.

Of the 176 *de novo* CNVs identified by our array, we confirmed 36, a true positive rate of 21%, similar to other whole-genome research CMA studies in which sensitivity is emphasized over specificity.¹⁶ Because of the research nature of this project, the settings to identify a CNV were liberal in order to detect previously unrecognized ID loci or uncover novel intragenic causative CNVs, and therefore a high false positive rate is expected. In terms of the specificity of our design, because of our Canadian health care structure, not all patients run on our custom array were fortunate to also have a clinical array performed. Therefore, without analyzing all of our patients on clinical arrays, it is unclear if our array design would miss

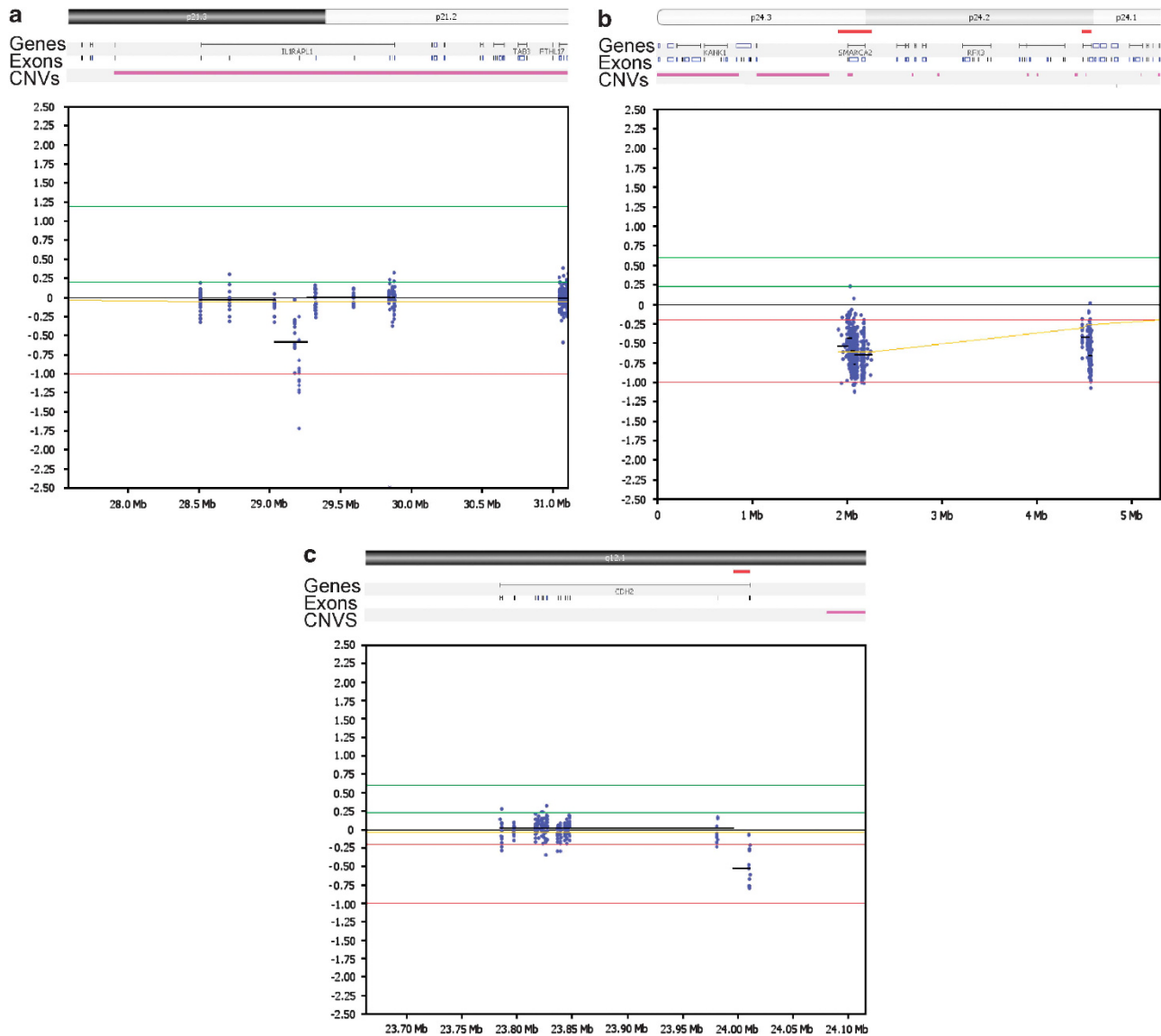


Figure 2 Examples of validated *de novo* CNVs. (a) 33Kb loss of exon 3 of the *IL1RAPL1* gene in patient 881. (b) A terminal loss of chromosome 9 in patient 412. (c) Loss of exon 1 of the *CDH2* gene in patient 576. In each image, the top portion shows the chromosome ideogram and relevant genes within the CNV. Losses are indicated by a red bar below the ideogram. The X-axis is base pairs and the Y-axis is the \log_2 ratio of the hybridization of the affected child relative to a normal parent. The red and green bars on either side of 0 designate the \log_2 ratio cutoffs for the program to flag a (heterozygous and homozygous) loss and gain, respectively.

potentially pathogenic CNVs. However, for patients for whom clinical CMA was performed, our design identified all pathogenic CNVs that were detected by clinical CMA. Only in one case (patient 412), our array missed one of the two CNVs identified on clinical CMA as discussed previously.

We identified 17 CNVs in 11 genes (five are known ID genes: *ARID1B*, *MEF2C*, *CHD7*, *UBE2A*, and *JMJD1C*) with epigenetic regulatory function and eight CNVs in six genes (five are known ID genes: *GRIK2*, *SHANK3*, *STXBPI*, *IL1RAPL1*, and *NRXN1*) with synaptogenic function. Many of the candidates we identified have not been previously reported to cause ID or have only been reported in an animal model or a few limited cases reports and require further study, beyond the scope of this work, to determine pathogenicity.

Although our custom microarray provided an effective platform for identifying known and novel CNVs associated with ID, its design has

several limitations: First, the lack of coverage outside our targeted regions usually prevented us from determining the CNV breakpoints. In our experience, as is also generally found,^{14,31} breakpoints defined by CMA are often imprecise, as they are based on statistical inference. Nevertheless, the location of breakpoints *within* exons on our custom chip is more precise because of the density of our probe coverage within targeted regions. The fact that breakpoints within an intronic sequence cannot be localized precisely should not alter our interpretation of an intragenic CNVs as long as the canonical splice donor and acceptor sequences are intact. However it is possible more complex rearrangements are present (for example, the multi-copy gains may not be positioned in tandem or even on the same chromosome) that we are unable to assess without breakpoint sequencing. We did not perform breakpoint sequencing as it is beyond the scope of this work, but doing so would have allowed us to

define the genotype more accurately and also infer genomic mechanisms for CNV causation. Second, the sparse and irregular genomic coverage prevents us from knowing whether pathogenic CNVs of untested regions are present in these patients. Both of these two issues could be resolved by adding a 'backbone' of regularly spaced probes throughout the genome. CMA is imprecise at estimating the number of copies present of a particular loci. qPCR was performed on all members of the trio and compared with a pooled reference set allowing us to resolve *de novo* CNVs calls within polymorphic loci (ie, a CNV called as a loss in a child can actually be a gain in the parent (or vice versa)). Such loci accounted for many of the false positive *de novo* CNVs in our CMA analysis. However, complex loci that have polymorphic alleles with several different copy-numbers or occur in various different overlapping sizes may have confounded both our CMA and qPCR analysis. Finally it must be borne in mind that in those cases where a clear genotype-phenotype correlation has not been established, the rare *de novo* event we have detected may not be necessary and sufficient to produce ID in the affected child. This is almost certainly true of the patients harboring CNVs involving *JARID2* exon 6. As information from higher resolution CMA and whole-exome and -genome sequencing studies of well-phenotyped ID patients accumulate, it should be possible to characterize the pathogenicity of many of the CNVs we found that are currently of uncertain clinical significance.

In summary, CMA was performed on 165 idiopathic ID trios using an exon-level-resolution custom microarray and *de novo* CNVs were confirmed in 32 trios. Sixty-four percent of our validated *de novo* CNVs were intragenic, including three pathogenic CNVs not identified by clinical CMA. Intragenic CNVs are likely a significant contributor to genetic causes of ID and will be missed with current commercially available clinical arrays using current clinical CMA guidelines.³¹

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We gratefully acknowledge all participating families. We thank Elaine Chan and Eric Zhao for their assistance. FRZ was supported by a Killam Doctoral Scholarship and Child and Family Research Institute Fellowship. This work was supported by the Canadian Institute of Health Research (grant number MOP 82828).

- Friedman J, Adam S, Arbour L *et al*: Detection of pathogenic copy number variants in children with idiopathic intellectual disability using 500K SNP array genomic hybridization. *BMC Genomics* 2009; **10**: 526.
- Jaillard S, Drunat S, Bendavid C *et al*: Identification of gene copy number variations in patients with mental retardation using array-CGH: Novel syndromes in a large French series. *Eur J Med Genet* 2010; **53**: 66–75.
- 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.
- Inlow JK, Restifo LL: Molecular and comparative genetics of mental retardation. *Genetics* 2004; **166**: 835–881.
- Chelly J, Khelifaoui M, Francis F, Cherif B, Bienvenu T: Genetics and pathophysiology of mental retardation. *Eur J Hum Genet* 2006; **14**: 701–713.
- Ropers HH: Genetics of early onset cognitive impairment. *Annu Rev Genomics Hum Genet* 2010; **11**: 161–187.
- Tayeh MK, Chin EL, Miller VR, Bean LJ, Coffey B, Hegde M: Targeted comparative genomic hybridization array for the detection of single- and multiexon gene deletions and duplications. *Genet Med* 2009; **11**: 232–240.
- Dhami P, Coffey AJ, Abbs S *et al*: Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am J Hum Genet* 2005; **76**: 750–762.
- Wong LJ, Dimmock D, Geraghty MT *et al*: Utility of oligonucleotide array-based comparative genomic hybridization for detection of target gene deletions. *Clin Chem* 2008; **54**: 1141–1148.
- Saillour Y, Cossee M, Leturcq F *et al*: Detection of exonic copy-number changes using a highly efficient oligonucleotide-based comparative genomic hybridization-array method. *Hum Mutat* 2008; **29**: 1083–1090.
- Hamdan FF, Daoud H, Piton A *et al*: *De novo* SYNGAP1 mutations in nonsyndromic intellectual disability and autism. *Biol Psychiatry* 2011; **69**: 898–901.
- Hamdan FF, Gauthier J, Araki Y *et al*: Excess of *de novo* deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am J Hum Genet* 2011; **88**: 306–316.
- Griffith M, Tang MJ, Griffith OL *et al*: ALEXA: a microarray design platform for alternative expression analysis. *Nat Methods* 2008; **5**: 118.
- Lee C, Scherer SW: The clinical context of copy number variation in the human genome. *Expert Rev Mol Med* 2010; **12**: e8.
- Shaffer LG, Kashork CD, Saleki R *et al*: Targeted genomic microarray analysis for identification of chromosome abnormalities in 1500 consecutive clinical cases. *J Pediatr* 2006; **149**: 98–102.
- Yu S, Kiehl M, Stegner AL, Kibiriyeva N, Bittel DC, Cooley LD: Quantitative real-time polymerase chain reaction for the verification of genomic imbalances detected by microarray-based comparative genomic hybridization. *Genet Test Mol Biomarkers* 2009; **13**: 751–760.
- Davuluri RV: Application of FirstEF to find promoters and first exons in the human genome. *Curr Protoc Bioinformatics* 2003; **Chapter 4**: Unit 4.7.
- Hamdan FF, Gauthier J, Dobrzaniecka S *et al*: Intellectual disability without epilepsy associated with STXBP1 disruption. *Eur J Hum Genet* 2011; **19**: 607–609.
- Franek KJ, Butler J, Johnson J *et al*: Deletion of the immunoglobulin domain of IL1RAPL1 results in nonsyndromic X-linked intellectual disability associated with behavioral problems and mild dysmorphism. *Am J Med Genet A* 2011; **155A**: 1109–1114.
- Behnecke A, Hinderhofer K, Bartsch O *et al*: Intragenic deletions of IL1RAPL1: Report of two cases and review of the literature. *Am J Med Genet A* 2011; **155A**: 372–379.
- de Leeuw N, Bulk S, Green A *et al*: UBE2A deficiency syndrome: Mild to severe intellectual disability accompanied by seizures, absent speech, urogenital, and skin anomalies in male patients. *Am J Med Genet A* 2010; **152A**: 3084–3090.
- Nascimento RM, Otto PA, de Brouwer AP, Vianna-Morgante AM: UBE2A, which encodes a ubiquitin-conjugating enzyme, is mutated in a novel X-linked mental retardation syndrome. *Am J Hum Genet* 2006; **79**: 549–555.
- Zahir FR, Baross A, Delaney AD *et al*: A patient with vertebral, cognitive and behavioural abnormalities and a *de novo* deletion of NRXN1alpha. *J Med Genet* 2008; **45**: 239–243.
- Ching MS, Shen Y, Tan WH *et al*: Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B**: 937–947.
- Le Meur N, Holder-Espinasse M, Jaillard S *et al*: MEF2C haploinsufficiency caused by either microdeletion of the 5q14.3 region or mutation is responsible for severe mental retardation with stereotypic movements, epilepsy and/or cerebral malformations. *J Med Genet* 2010; **47**: 22–29.
- Visser LE, van Ravenswaaij CM, Admiraal R *et al*: Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nat Genet* 2004; **36**: 955–957.
- Lehman AM, Friedman JM, Chai D *et al*: A characteristic syndrome associated with microduplication of 8q12, inclusive of CHD7. *Eur J Med Genet* 2009; **52**: 436–439.
- Sharp AJ, Selzer RR, Veltman JA *et al*: Characterization of a recurrent 15q24 microdeletion syndrome. *Hum Mol Genet* 2007; **16**: 567–572.
- Haverfield EV, Whited AJ, Petras KS, Dobyns WB, Das S: Intragenic deletions and duplications of the LIS1 and DCX genes: a major disease-causing mechanism in lissencephaly and subcortical band heterotopia. *Eur J Hum Genet* 2009; **17**: 911–918.
- Tucker T, Montpetit A, Chai D *et al*: Comparison of genome-wide array genomic hybridization platforms for the detection of copy number variants in idiopathic mental retardation. *BMC Med Genomics* 2011; **4**: 25.
- 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.
- 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.
- Bailey JA, Kidd JM, Eichler EE: Human copy number polymorphic genes. *Cytogenet Genome Res* 2008; **123**: 234–243.
- Wisniewicka-Kowalnik B, Kastory-Bronowska M, Bartnik M *et al*: Application of custom-designed oligonucleotide array CGH in 145 patients with autistic spectrum disorders. *Eur J Hum Genet* 2012; **21**: 620–625.
- Boone PM, Bacino CA, Shaw CA *et al*: Detection of clinically relevant exonic copy-number changes by array CGH. *Hum Mutat* 2010; **31**: 1326–1342.

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