npg

ARTICLE

High-resolution SNP arrays in mental retardation diagnostics: how much do we gain?

Laura Bernardini[1], Viola Alesi[1,2], Sara Loddo[1,2], Antonio Novelli[1], Irene Bottillo[1], Agatino Battaglia[3], Maria Cristina Digilio[4], Giuseppe Zampino[5], Adam Ertel[2], Paolo Fortina*[,2,6], Saul Surrey[7] and Bruno Dallapiccola[1]

We used Affymetrix 6.0 GeneChip SNP arrays to characterize copy number variations (CNVs) in a cohort of 70 patients previously characterized on lower-density oligonucleotide arrays affected by idiopathic mental retardation and dysmorphic features. The SNP array platform includes ~900 000 SNP probes and 900 000 non-SNP oligonucleotide probes at an average distance of 0.7 Kb, which facilitates coverage of the whole genome, including coding and noncoding regions. The high density of probes is critical for detecting small CNVs, but it can lead to data interpretation problems. To reduce the number of false positives, parameters were set to consider only imbalances >75 Kb encompassing at least 80 probe sets. The higher resolution of the SNP array platform confirmed the increased ability to detect small CNVs, although more than 80% of these CNVs overlapped to copy number 'neutral' polymorphism regions and 4.4% of them did not contain known genes. In our cohort of 70 patients, of the 51 previously evaluated as 'normal' on the Agilent 44K array, the SNP array platform disclosed six additional CNV changes, including three in three patients, which may be pathogenic. This suggests that about 6% of individuals classified as 'normal' using the lower-density oligonucleotide array could be found to be affected by a genomic disorder when evaluated with the higher-density microarray platforms.

INTRODUCTION

Copy number variations (CNVs) are defined as DNA changes >1 Kb that differ among individuals. They include duplications or deletions and can involve large DNA regions.[1–3] CNVs have an important role in modulating the phenotypic spectrum both in single gene and multigenic diseases.[3] They are not randomly distributed across the genome, and their presence correlates with chromosomal structural features (eg, inverted repeats, duplications, highly homologous regions and so on).[2,4–8]

In the past few years, DNA microarrays have increased capabilities for the detection of both pathological and neutral polymorphic variations by analyzing the entire genome with a resolution depending on the number, length and type of probes.[1,9–14] Genome tiling microarrays are now available, allowing DNA analysis at an intermediary resolution between cytogenetic techniques (≥5 Mb) and conventional DNA sequence analysis (1–800 bp).[9]

Mental retardation (MR) is a relatively frequent disorder affecting about 3% of the general population. MR etiological diagnoses are made usually in fewer than half of the affected individuals.[15] Genetic causes account for 17.4–47.1% of cases, with reported frequencies also varying on the basis of the different techniques used for analysis.[15,16] Several years ago, microarray platforms were used for the first time in clinical genetics, showing that abnormalities in genomic copy number

account for about 10–15% of patients affected by MR, especially when it is associated with multiple congenital anomalies (MCA) and/or dysmorphism.[17] Microarrays represent a robust and high-resolution method of analysis. The recent increase in various commercially available microarray platforms encouraged their application in diagnostic settings, and it is expected to improve procedures for the establishment of genetically determined MR etiology.[18] Among commercial platforms, SNP arrays offer the highest resolution using approximately two million probes. These arrays were originally designed for genome-wide genotyping, but they can also be used to detect CNV.[19] This approach involves only one DNA sample to be hybridized on the chip. Software is then used to reveal copy number at each locus quantifying the fluorescent signal intensity and comparing it with HapMap controls (International HapMap Project, 2005; http://www.hapmap.org/). One limitation of these arrays is that probe selection is not homogeneous, as it reflects the distribution of SNPs in the genome, and SNPs are very sparse in genomic segments containing low-copy repeats and segmental duplications.[9] SNP arrays frequently offer a better detection of CNVs in gene-poor regions.[20] To overcome this restriction, the latest generation high-resolution SNP array platforms were generated by adding nonpolymorphic sequences to cover the SNP-poor region (Illumina, San Diego, CA, USA and Affymetrix, Santa Clara, CA, USA). For example, the Affymetrix 6.0

[1]CSS Hospital, IRCSS, San Giovanni Rotondo and CSS-Mendel Institute, Rome, Italy; [2]Department of Cancer Biology, Kimmel Cancer Center, Thomas Jefferson University, Jefferson Medical College, Philadelphia, PA, USA; [3]Stella Maris Clinical Research Institute for Child and Adolescent Neurology and Psychiatry, Pisa, Italy; [4]Medical Genetics, IRCCS Bambino Gesù Pediatric Hospital, Rome, Italy; [5]Department of Pediatrics, Cattolica University, Rome, Italy; [6]Department of Experimental Medicine, Sapienza University, Rome, Italy; [7]Cardeza Foundation for Hematologic Research, Department of Medicine, Jefferson Medical College, Thomas Jefferson University, Philadelphia, PA, USA
*Correspondence: Dr P Fortina, Kimmel Cancer Center, Thomas Jefferson University, 233 South 10th Street, 1009-J BLSB, Philadelphia, PA 19107, USA.
Tel: +1 215 955 0683; Fax: +1 215 503 9142; E-mail: paolo.fortina@jefferson.edu

GeneChip consists of both 906 600 SNP sequences and about 900 000 nonpolymorphic oligonucleotides, with an overall spacing of 1 sequence per 700 bp along the entire genome.[21] The theoretical resolution of a microarray depends on the length, number and genomic density of probes; however, many other parameters can determine the practical resolution, including the level of experiment noise and the sensitivity of copy number measurements.[9,22,23] Thus, in oligonucleotide arrays as in SNP arrays, the reduced length of probes tends to generate a lower signal-to-noise ratio of hybridization, and analyses need to average data from several consecutive spotted sequences, thus diminishing the overall resolution.[9]

In this study, we analyzed 70 karyotyped patients affected by MR/MCA, who were previously analyzed with an oligonucleotide array with an average spacing of 32 kb. CNVs detected by SNP array were selected following the workflow as previously suggested,[24] and they were verified systematically by other techniques (fluorescence in situ hybridization (FISH) and/or quantitative-PCR (qPCR)). The purpose of this study was to assess Affymetrix GeneChip 6.0 array sensitivity in CNV detection to establish the effective resolution of this platform and to explore the role of CNVs in the definition of as-yet-unexplained MR patients. In particular, we intended to evaluate the potential diagnostic usefulness of this technique coupling a more sensitive chip with strict clinical criteria.

MATERIALS AND METHODS

Patients
A total of 70 patients were evaluated. They were selected from the cohort of subjects referred to our diagnostic laboratory for CNVs analysis in the 2006–2007 2-year period. All individuals were karyotyped (>450 band resolution): 64 had a normal karyotype, 5 had apparently balanced structural rearrangements and 1 patient had a subtelomeric rearrangement detected by FISH analysis. Diagnostic screening was performed using an oligonucleotide-based array (44K – AMADID 014950; Agilent Technologies, Walldbron, Germany). A total of 19 patients were positive in this initial screening. This group of individuals (SET I) was used for validation purposes. The remaining 51 patients (SET II) were analyzed to assess whether the higher resolution of GeneChip 6.0 resulted in an effective higher sensitivity in detecting smaller pathogenic CNVs in these individuals. These two platforms were comparable in terms of the cost and timetable of a single experiment. All studies were approved by the Institutional Review Board committee and conducted with patients' informed consent. All patients were affected by moderate-to-severe MR (QI<50), showed dysmorphic features and/or developmental delay and/or had at least one major severe congenital malformation and/or multiple mild congenital anomalies. A commercially available NucleoSpin Blood kit (Macherey-Nagel GmbH & Co, Düren, Germany) was used to extract genomic DNA from 2 ml of peripheral blood.

Controls
A data set of genotype and copy number calls on 270 control samples from the International HapMap Project was used (http://www.hapmap.org/). The SNP 6.0 copy number calls from these HapMap samples were provided by Affymetrix as a reference model file for use in Genotyping Console software (Affymetrix).

Affymetrix GeneChip mapping 6.0 microarray
The GeneChip 6.0 platform consists of about 906 600 SNP sequences and about 900 000 nonpolymorphic oligonucleotides, which cover the whole genome with an average spacing of 0.7 Kb. In two separate reactions, genomic DNA (250 ng) was digested with NspI or StyI (New England Biolabs, Ipswich, MA, USA), as recommended by the manufacturer (Affymetrix). After digestion, an adaptor was linked to the restricted fragments, the reaction was diluted 4× and the fragments were amplified by PCR. After purification using Magnetic Beads (Agencourt Bioscience Corporation, Beverly, MA, USA), 90 µg of PCR products was fragmented and end labeled using 30 U/µl of terminal deoxynucleotidyl

transferase, and then hybridized for 16–18 h to the Affymetrix 6.0 chip at 49°C in a GeneChip Hybridization Oven 640 (Affymetrix). The chips were washed, stained in GeneChip Fluidic Station 450 (Affymetrix) and scanned with Scanner 3000 7G (Affymetrix).

Genotype and copy number calls
Affymetrix CEL files were loaded into Genotyping Console (GTC) 3.0.1 for genotype and copy number analysis. The GTC Contrast Quality Control (QC) metric was used to filter out low-quality samples using the default threshold of 0.4. Copy number analysis, using the Affymetrix HapMap270 Reference Model File for comparison, was performed with the default settings in GTC, including regional GC correction. Detailed descriptions of the algorithms implemented by GTC copy number analysis are provided in Korn et al.[25] Briefly, GTC applies the BRLMM-P Plus algorithm, which uses a one-dimensional Gaussian mixture model to establish the intensity of each genotype and copy number probe at discrete levels relative to the reference model file. A Hidden Markov Model is then applied for smoothing the copy number data on the basis of neighboring values in a region.

Copy number segment analysis
The GTC segment-reporting tool was supplied to identify regions in the genome with a default setting of at least 25 markers showing consensus for gain or loss spanning at least a 75-kb region. Data for regions meeting these criteria were exported for further analysis in Microsoft Excel. These results of the segment analysis were used to compare aberrant regions against known copy number polymorphisms and select copy variations for validation. The copy number aberrations were further filtered to select CNVs more likely to be pathogenic on the basis of the size and number of markers in the region. A CNV was scored when it encompassed >75 Kb and included at least 80 markers. Considering that the average spacing of the targets of this array is 0.7 Kb, one would expect roughly 107 SNP markers ((1 SNP/0.7 kb)×75 kb= 107 SNPs). The smallest CNVs were considered only when at least 75% of the CNV length was covered by markers.

Validation of CNVs
The database of Genomic Variants (http://projects.tcag.ca/variation/), UCSC Genome browser (http://www.genome.ucsc.edu/) and DECIPHER (http://www.sanger.ac.uk/PostGenomics/decipher), were used to identify known genomic variants, genes and MR-associated regions. Each CNV containing known genes and/or miRNA sequences not found in controls was attempted to be confirmed by other methods. FISH analysis was used to confirm deletions >150 Kb and duplications >400 Kb, whereas custom quantitative-PCR assay was used to confirm smaller CNVs.

FISH
To confirm new imbalances, appropriate BACs (bacterial artificial chromosomes) were selected from a genome-wide 32K BAC library (BACPAC Resources Center, Children's Hospital Oakland Research Institute, Oakland, CA, USA) (http://www.chori.org/bacpac). Probes were prepared and FISH was performed as previously described.[26] An Eclipse 80i fluorescence microscope equipped with a computerized system (Genikon; Nikon, Florence, Italy) was used to analyze and acquire images. To confirm an imbalance, about 30 metaphase spreads and/or nuclei for each sample were counted.

qPCR
Smaller sized DNA copy number changes were confirmed using an AB 7000 Sequence Detection System (Applied Biosystem, Foster City, CA, USA) and DNA-binding dye SYBR Green (Invitrogen Corporation, Carlsbad, CA, USA) as described by Carbone et al.[27]

X-inactivation assay
The X-chromosome inactivation pattern in patient 29 was assessed by studying the methylation state of the androgen receptor gene.[28] PCR products were separated on an ABI Prism 3100 DNA sequencer and analyzed using GeneScan

version 3.5.1 (Applied Biosystems) and Genotyper software version 2.1.5 (Applied Biosystems).

RESULTS

This study focused on testing for CNVs in patients with MR on a high-resolution SNP array platform consisting of about 1 800 000 probes including polymorphic and nonpolymorphic sequences. A total of 70 MR/MCA patients were hybridized on a GeneChip 6.0 platform and data were analyzed using Genotyping Console software (GTC v3.0.1). SNP array analysis detected the presence of CNVs in all patients with a highly variable number ranging from 10 to 313 per subject.

Using filters described in Materials and Methods, an average of 21 CNVs were detected in each sample (range 8–41), with a size range of 0.76–16.314 Mb. This number was reduced further, following the recommendations of Koolen et al.[24] CNVs were not considered as pathogenic if the rate of overlap with known nonpathogenic polymorphic CNVs was >50–70% (81% of CNVs and 100% of CNVs smaller than 100 kb), or if the CNV did not involve known genes (4.4% of nonpolymorphic CNVs). CNVs with a probe density <2 kb were excluded as well. These rules were not considered for CNVs affecting regions known to be involved in microdeletion/microduplication syndromes and for X-linked CNVs occurring in male patients. All selected CNVs were confirmed using other techniques and, when possible, this test was expanded to parents to distinguish between de novo-generated and inherited CNVs. Clinical features, SNP array results, standard karyotypes and array-CGH data of patients are summarized in Table 1. The number of probes present in each affected chromosome region, their density, degree of overlap with known common CNV regions, confirmation technique and CNV origin are also listed.

SET I

This set included the 19 patients previously shown to be positive for copy number changes using oligonucleotide array analysis. SNP array analysis confirmed the presence of all CNVs previously detected: 14 deletions ranging in size from 0.257 to 6.494 Mb and 9 duplications of 0.583–7.930 Mb. In total, seven patients (1, 4, 7, 8, 13, 16 and 18) displayed a pathogenic deletion, including the subtelomeric deletion of 1p, 2q and 9q. A total of five patients (2, 9, 10, 12, 15) had a duplication including the MECP2 duplication, and duplication of the proximal 15q region associated with autism in two subjects. Patient 6 was shown to be the carrier of two losses on both analyses: deletion of 4.9 Mb at 3q23q24, near the break point of a de novo reciprocal translocation (3; 21), and deletion of about 3.2 Mb at 11p11.2. Patient 11 had a terminal deletion and associated duplication of two contiguous regions at 10pter, whereas Patient 17 harbored a terminal deletion associated with a contiguous duplication on chromosome 1 (Figure 1).

In four patients (3, 5, 13 and 19), SNP array analysis showed additional CNVs not detected by the oligonucleotide array. Patient 3 had a 1.4-Mb deletion of 3p25.3p25.2 on both platforms, whereas the SNP array disclosed an additional 0.871-Mb duplication of 18q22.1, which proved to be maternally inherited by FISH analysis. Patient 5 disclosed the duplication of two separate but neighboring regions (3.494 and 6.091 Mb) on chromosome 5q, confirmed by FISH (Figure 2a and b). These duplications were shown to derive from the malsegregation of a complex balanced paternal rearrangement involving chromosomes 3, 5 and 11 (Figure 2c). SNP arrays also showed a duplication of 0.140 Mb at Xq13.1, considered to be a false-positive result, as qPCR performed on PJA1 disclosed only two copies

similar to that in controls. Patient 13 showed a 0.373-Mb subterminal deletion at 9q34.3 on both analyses, in addition to a paternally inherited gain of 0.684 Mb at 1q43 (ZP4 gene) on SNP arrays. Patient 19 showed a de novo gain of 0.188 Mb at 5q14.3, encompassing the CETN3 gene, in addition to a deletion at 1p35.3p35.2 disclosed by both platforms.

SET II

This set included 51 samples previously found to be negative for copy number changes using oligonucleotide arrays. A total of 48 patients showed no CNV change after SNP array analysis. Three samples (29, 54 and 66) were positive on SNP arrays, disclosing CNVs ranging in size from 0.352 to 0.583 Mb. In patient 29, the SNP array showed a duplication of 0.583 Mb at Xp11.4 containing two genes (TSPAN7 and MID1IP1) (Figure 3a). Although the parents' DNA was not available to assess the origin of this imbalance, an X-inactivation assay showed a random inactivation of X chromosomes (68%) (Figure 3a). Patient 54 showed a deletion of 0.545 Mb at 9q33.1 involving TRIM32 and ASTN2 (Figure 3b), which was not present in his mother, whereas the father's DNA was not available to elucidate the origin of CNV. In patient 66, the SNP array disclosed a de novo duplicated region of about 0.52 Mb containing three known genes (ANKRD56, SEPT11 and CCNI) at 4q21.1 (Figure 3c).

DISCUSSION

Microarray-based platforms to be applied in diagnostic practice require a high level of robustness, reproducibility and effective resolution in detecting CNVs. Hehir-Kwa et al, performed a statistical power analysis to compare four different genomic platforms and showed that the actual performance of a microarray depends not only on the spacing of probes but also on the noise and sensitivity of copy number variation calls. In particular, the number of consecutively spotted sequences needed to detect a reliable CNV affects the resolution power.[22] Coe et al,[29] introduced the term 'functional resolution' to indicate the actual performance of a microarray-based platform, which is also estimated considering the uniformity of element spacing on the array and the sensitivity of each platform to single-copy alterations.

In our study, the high-resolution SNP array GeneChip 6.0 (Affymetrix) sensitivity in identifying copy number changes was tested by analyzing 19 patients previously shown positive and 51 patients who were negative to copy number analyses on Agilent lower-resolution arrays. Data were processed using selection criteria, and CNVs smaller than 75 kb and represented by less than 80 targets were excluded. These parameters are appropriate, but may become less rigorous with the decreasing size of imbalances. The parameters used could be useful to screen patients with phenotypes suggestive of genomic imbalance, although they may very likely underestimate the detection of very small CNVs. Using this approach, we found only one false-positive CNV mapping to Xq13.1, spanning 0.140 Mb, which was not confirmed by other techniques (patient 5; Table 1). Moreover, small CNVs with an average length of several Kbs cover about 12% of the normal human genome,[11] and it is difficult to assess whether they have any clinical consequence.[30] To optimize the management of CNV analyses, a workflow for clinical practice was published by Koolen et al.[24] In our study, these guidelines were followed to further select the most likely pathogenic CNVs, and it was shown that more than 80% of CNVs detected by high-resolution platforms were polymorphic, nonpathogenic copy number changes. In particular, we found that 100% of CNVs showing a length lower than 100 Kb completely overlapped to polymorphic genomic regions. No false

Table 1 Clinical and molecular features of patients with positive SNP array analysis

| | Patients | | | Array-CGH analysis | | SNP array analysis (segment reporting tool data) | | | | | | | | |
No.	Clinical features	Age	Sex	Standard karyotype	Result	CNV	Cytoband	Size (kb)	Probes	Distance (kb)	%	Start	End	Origin
I SET														
1	MR, dysmorphisms	7y	F	46,XX	del(5)(q14.2q14.2)	Loss	5q14.2	257	136	2	0	82 399 900	82 657 181	N.A.
2	Moderate MR, severe hypotonia, spasticity, white matter anomalies	7y	M	46,XY	dup(X)(q28q28)	Gain	Xq28	731	318	2	100	152 680 476	153 436 362	de novo
3	MR, speech delay, stereotyped behavior, mild dysmorphisms	5y, 9m	F	46,XX	del(3)(p25.3p25.2)	Loss	3p25.3p25.2	1395	973	1	2	10 882 556	12 277 462	de novo
4	Moderate MR, dysmorphisms, scoliosis, obesity, precocious puberty	16y	F	46,XX	Negative / del(14)(q32.2q32.1)	Gain / Loss	18q22.1 / 14q32.2q32.31	871 / 1105	684 / 585	1 / 2	67 / 26	63 801 893 / 99 465 564	64 673 225 / 100 570 110	Inherited mat / de novo
5	Moderate MR, microcephaly, dysmorphisms epilepsy	1y, 5m	F	46,XX,t(3;11)(q25.3;q23.3)	dup(5)(q31.1q31.3)	Gain	5q31.1q31.3	5798	3187	2	50	134 771 280	140 863 043	de novo
6	Moderate MR, hypotonia, congenital hypothyroidism	1y, 2m	F	46,XX,t(3;21)(q24;q22)	dup(5)(q34) / Negative / del(3)(q23q24)	Gain / Gain / Loss	5q34 / Xq13.1 / 3q23q24	3486 / 140 / 4865	2183 / 101 / 3137	2 / 1 / 2	6 / 0 / 30	161 764 915 / 68 314 964 / 143 887 022	165 259 694 / 68 455 333 / 148 751 955	de novo / / / de novo
7	MR, dysmorphisms	1y, 6m	F	46,XX	del(11)(p11.2p11.2)	Loss	11p11.2	3141	1497	2	10	45 366 618	48 611 419	de novo
8	Moderate psychomotor and speech delay, dysmorphisms	3y, 10 m	M	46,XY	del(1)(q23.3q24.2) / del(11)(q13.2q13.2)	Loss / Loss	1q23.3q24.2 / 11q13.2	5894 / 435	4608 / 184	1 / 2	22 / 1	160 417 164 / 67 647 865	166 311 507 / 68 083 239	de novo / de novo
9	Moderate MR, dysmorphisms, epilepsy	2y, 1m	M	46,XY	dup(X)(p22.31p22.31)	Gain	Xp22.31	583	474	1	100	7 522 572	8 105 356	N.A.
10	MR, dysmorphisms	4y, 3m	M	46,XY	dup(15)(q11.2q13.2)	Gain	15q11.2q13.2	7930	4566	2	58	20 307 305	28 237 649	de novo
11	Moderate MR, dysmorphisms, cleft lip and palate, ToF, hypoparathyroidism	2y, 7m	F	46,XX	del(10)(p15.3p15.1)	Loss	10p15.3p15.1	6494	5543	1	26	62 797	6 567 350	de novo
12	Moderate MR, mild dysmorphisms, hypotonia, joint laxity	2y, 6m	M	46,XY	dup(10)(p15.1p14) / dup(X)(p22.31)	Gain / Gain	10p15.1p14 / Xp22.31	2223 / 1653	2006 / 1225	1 / 1	21 / 100	6 568 384 / 6 459 754	8 791 078 / 8 113 190	de novo / Inherited mat
13	Moderate MR, dysmorphisms, AVCD	1y, 1m	F	46,XX	del(9)(q34.3q34.3) / Negative	Loss / Gain	3q34.3 / 1q43	373 / 684	174 / 560	2 / 1	18 / 46	139 510 436 / 235 646 856	139 883 030 / 236 331 048	de novo / Inherited pat
14	Moderate MR, dysmorphisms, feeding difficulties, rhizomelic limb shortening, axial hypotonia, bilateral hypoacusia	3y, 6m	M	46,XY	dup(X)(q12q21.1)	Gain	Xq12q21.1	16314	7645	2	22	66 352 026	82 814 697	de novo
15	Moderate MR, stereotypical language, autism, dysmorphisms, joint laxity, scoliosis	9y, 10m	M	46,XY	dup(15)(q11.2q13.1)	Gain	15q11.2q13.1	5033	3427	1	56	21 192 943	26 234 399	de novo
16	Developmental delay, severe MR, stereotyped behavior, dysmorphisms	4y, 6m	M	46,XY	del(18)(q21.2q21.2)	Loss	18q21.2	373	236	2	8	51 111 859	51 485 203	de novo
17	Moderate MR, dysmorphisms	3y, 9m	F	46,XX. ish del(1)(pter)	del(1)(p36.33p36.33)	Loss	1p36.33	1629	378	4	97	554 484	2 183 946	de novo
18	Moderate MR, dysmorphism	2y, 6m	M	46,XY	dup(1)(p36.33p36.32) / del(2)(q37.3q37.3)	Gain / Loss	1p36.33p36.32 / 2q37.3	1744 / 5135	695 / 2925	3 / 2	62 / 28	2 184 496 / 237 603 010	3 928 784 / 242 738 117	de novo / de novo
19	Severe MR, behavioral problem, speech delay, dysmorphisms, hypotonia, angiomas	6y, 6m	F	46,XX	del(1)(p35.3p35.2)	Loss	1p35.3p35.2	3621	2112	2	16	28 237 648	31 858 224	de novo
					Negative	Gain	5q14.3	188	107	2	0	89 601 661	89 789 670	de novo
II SET														
29	Psychomotor delay, dysmorphisms, cerebellar hypoplasia	8y	F	46,XX	Negative	Gain	Xp11.4	583	421	1	22	38 310 495	38 893 481	N.A.
54	Moderate MR, dysmorphisms, hypotonia, seizures, obesity, bilateral cryptorchidism, café-au-lait spots	10y	M	46,XY	Negative	Loss	9q33.1	545	613	1	22	118 489 051	119 034 042	de novo[a]
66	Moderate MR, dysmorphisms, ASD ostium secundum type, cryptorchidism	4y, 6m	M	46,XY	Negative	Gain	4q21.1	352	251	1	6	77 973 643	78 325 578	de novo

ASD, atrial septal defect; AVCD, atrioventricular conduction defect; CNV, copy number variations; F, female; M, male; m, months; MR, mental retardation; ToF, tetralogy of Fallot; y, years.
[a]Father's DNA not available.
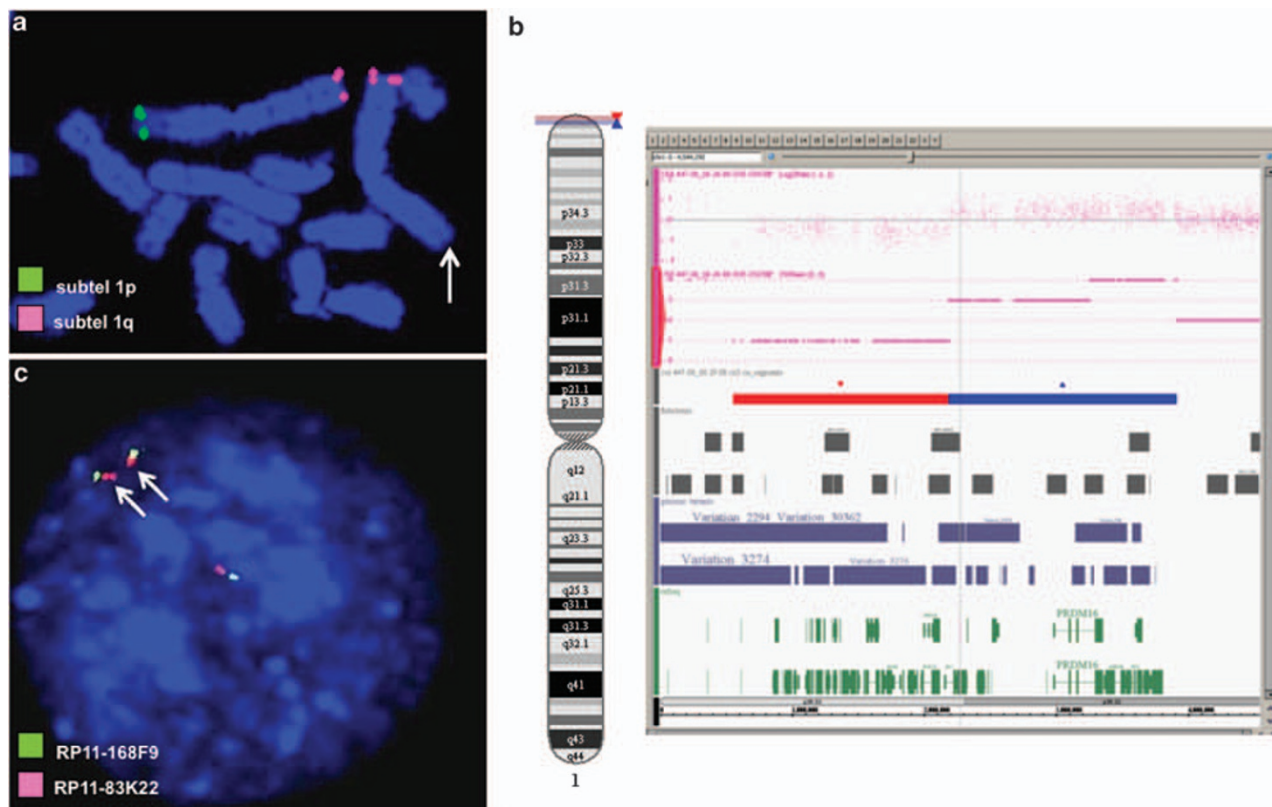In grey: CNV not confirmed by other techniques.

Figure 1 Patient 17: (**a**) Subtelomeric deletion of the 1p36.33 region disclosed by fluorescence *in situ* hybridization (FISH) analysis (arrow). (**b**) Single-nucleotide polymorphism (SNP) array analysis shows the deletion (red bar and arrow), followed by the duplication of a contiguous segment (blue bar and arrow). Although copy number segment analysis indicates a simple gain, the CN state (pink line) suggests that a part of the duplicated region could present a triplication, may be because of the high copy number variability of this region (**c**) FISH analysis on interphase nuclei shows the inverted duplication of the 1p36.33p36.32 region (arrows).

negatives were recorded in our study, as all previously detected CNVs were identified by SNP arrays (SET I). These CNVs were regarded as pathogenic, as they either corresponded to known regions involved in microdeletion/microduplication syndromes and/or were quite large imbalances encompassing several genes, suggesting a causal relationship with the phenotype. The imbalance size and gene content are important parameters to be considered while assessing the pathogenic role of CNVs.[24,31] Moreover, for 12 of our 19 positive patients it was possible to trace the CNV origin, and only 1 was found to be inherited (patient 12). This was a male patient hemizygous for an Xp22.31 duplication inherited from his unaffected mother. The duplication was because of a nonhomologous allelic recombination event among members of the *VCX/Y* gene family flanking the rearranged segment. The reciprocal deletion of this genomic segment is associated with 80–90% of X-linked icthyosis/MR cases[32] and this duplication was reported in a child with MR,[28] although its pathogenicity has not been definitely proved.[33]

Microarray analysis showed that chromosome anomalies detected by standard and molecular cytogenetic techniques may uncover more complex rearrangements resolved by CNV studies. For example, in patient 5, two duplications of about 6 and 3.5 Mb on chromosome 5q were detected. The patient was affected by developmental delay associated with dysmorphic features, and harbored an apparently balanced translocation involving chromosomes 3q and 11q. As the translocation was inherited from the unaffected father, in the beginning it was not considered to be related to the patient's phenotype.

Integrated microarray and FISH analyses showed that the patient was heterozygous for duplications resulting from a malsegregation of a more complex balanced paternal rearrangement, involving chromosomes 3, 5 and 11. In patient 17, deletion of the 1pter region was diagnosed by FISH-based subtelomeric investigation. However, the patient's clinical features only partly overlapped the phenotype of a 1p36 deletion syndrome,[34] prompting microarray analysis to further characterize the rearrangement. CNV analysis confirmed a 1.4-Mb deletion of the subtelomeric region of 1p36.33, which was smaller compared with the one usually associated with the 1p36.33 syndrome.[35] Array-based analysis also disclosed a 1.6-Mb duplication of a contiguous 1p36.32 segment (Figure 1). This duplication was reported previously in association with a 1p36 deletion,[36] and it is likely that its frequency has been underestimated lacking appropriate diagnostic techniques. Both the size of the deleted segment and the cooccurrence of the duplication could well explain the atypical clinical features of this patient, including distinct facial appearance and absent heart defects and seizures.[34] Interestingly, the Affymetrix SNP platform detected these imbalances, but the CNV overlapping rate indicated that these copy number changes were neutral polymorphisms (Table 1). Recent data suggest that the total genomic content of known common human CNVs in databases is overestimated.[37] Therefore, caution should be exercised while interpreting the significance of the CNV rate, especially for changes involving large genomic areas particularly rich in small polymorphic CNV segments.
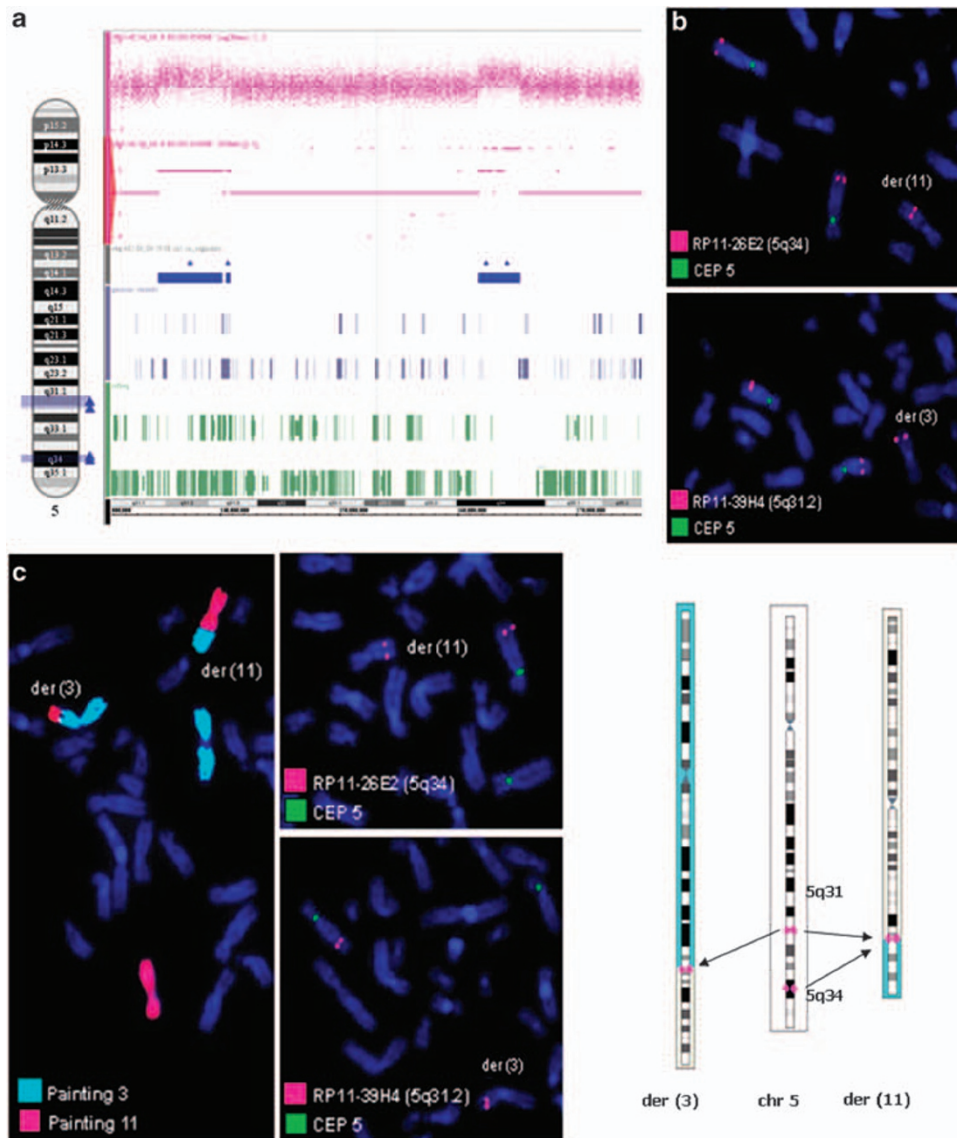
Figure 2 (a) Duplications at 5q31 and 5q34 (patient 5) identified by single-nucleotide polymorphism (SNP) arrays and indicated by blue bars and arrows. (b) Fluorescence *in situ* hybridization (FISH) analysis shows that duplicated segments map to chromosomes involved in a paternal translocation: in particular, the 5q31 region is partially translocated onto the derivative chromosome 3 (not shown) and is partially translocated onto the derivative chromosome 11 at 11q23.3 (top); 5q34 is translocated onto the derivative chromosome 3 at 3q25.3 (bottom). (c) Left: FISH analysis shows a complex translocation involving chromosomes 3, 5 and 11 in the patient's father as described in ideograms (right).

The increased resolution provided by SNP array platforms provides an increased ability to detect small CNVs. In our cohort of patients, six additional copy number changes were detected, including four imbalances, which may be considered pathogenic, but further studies are required to substantiate this contention. In particular, SNP arrays disclosed likely causative CNVs in three patients regarded as normal on the basis of previous CNV analyses. This means that ~6% of patients belonging to the present SET II were found to be affected by a 'genomic disorder'. In patient 29, a duplication at Xp11.4 involving the *TSPAN7* (OMIM 300096) gene was detected (Figure 3a). *TSPAN7* mutations have been associated with X-linked nonsyndromic MR, and a duplication of this gene has been reported in three patients,[38] although the pathogenic function of copy number gains remains controversial. In our case, parents were not available to assess the origin of these CNVs, but in structural abnormalities of the X

chromosome, females presenting with a random X-inactivation pattern is associated with a severe phenotype.[39] SNP array analysis in patient 54 disclosed a loss at 9q33.1 including *ASTN2* and *TRIM32* genes. *ASTN2* is highly expressed in the central nervous system and a relationship with schizophrenia was suggested recently.[40] *TRIM32* has been associated with the Bardet–Biedl syndrome 11 (BBS11; OMIM 209900). Considering that the inheritance of this disorder was hypothesized to be triallelic, it can be surmised that *TRIM32* heterozygotes contribute to the BBS phenotype, whereas a deletion of this gene can result in a distinct phenotype. In patient 66, 4q21.1 duplication encompassed the *ANKRD56*, *SEPT11* and *CCNI* genes. Parental FISH and qPCR analyses showed the *de novo* origin of CNV. However, the role of these genes is not clear at present, thus preventing the possibility of establishing a correlation between their overexpression and the patient's phenotype. Notably, the *SEPT11* gene was
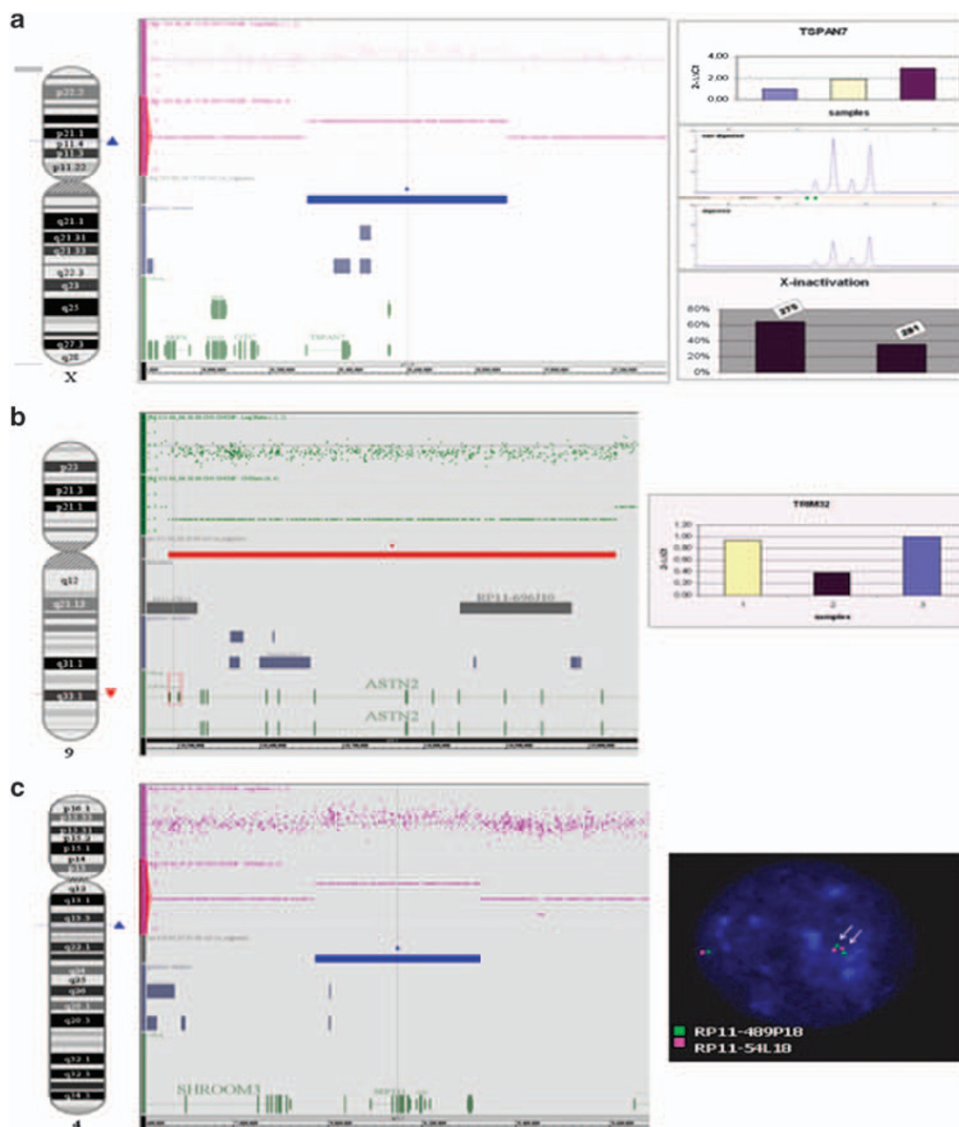
Figure 3 (**a**) Patient 29: gain at Xp11.4, 0.583 Mb in size, including the *TSPAN7* gene (blue bar and arrow). Top right: duplication was confirmed by qPCR on *TSPAN7* (blue and yellow bars are male and female controls, respectively, whereas the violet bar represents the patient). Center and below: random X inactivation was shown on the androgen receptor gene (AR), showing two alleles (275 and 281), on the basis of the ratio of areas of two alleles after digestion. (**b**) Patient 54: deletion of about 0.5 Mb at 9q33.1 (red bar and arrow), including *ASTN* and *TRIM32* genes. *TRIM32* is indicated by the red square, as this gene is included in the *ASTN* sequence. Deletion was confirmed by qPCR on *TRIM32* (blue and yellow bars are controls, whereas the violet bar represents the patient). (**c**) Patient 66: ~0.3 Mb duplication at 4q21.1 (blue bar and arrow) detected by single-nucleotide polymorphism (SNP) array and confirmed by interphase fluorescence *in situ* hybridization (FISH) (arrows).

shown to be expressed in CNS, both in cell bodies and dendrites,[41,42] and septins are GTP-binding proteins interacting with developing myelin structures.[43]

No mosaic genomic alteration was found in our cohort of patients, either indicating that mosaic imbalances are a very rare cause of MR/MCA or reflecting the low sensitivity of the SNP array platform to detect CNVs in samples containing different cell populations. This point was discussed by Zhang et al,[44] who compared different high-resolution microarrays and observed that some platforms, including the Affymetrix SNP array, could escape the detection of mosaic cases because of data noise.

In summary, genomic microarrays are important tools for studying patients with idiopathic MR. The opportunity of identifying very small imbalances using high-resolution microarray platforms improves our ability to detect the genetic defects responsible for rare forms of MR. However, this opportunity should be weighed against the detection of false positives and small CNVs devoid of any obvious pathogenic effect, including CNVs inherited from unaffected parents. Therefore, stringent parameters must be used for data analysis, and parents' DNA should be available whenever these genomic platforms are used for diagnostic purposes.

1 Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. *Nat Rev Genet* 2006; **7**: 85–97.
2 Chance PF, Abbas N, Lensch MW *et al*: Two autosomal dominant neuropathies result from reciprocal DNA duplication/deletion of a region on chromosome 17. *Hum Mol Genet* 1994; **3**: 223–228.
3 Stankiewicz P, Lupski JR: Genome architecture, rearrangements and genomic disorders. *Trends Genet* 2002; **18**: 74–82.
4 Cooper GM, Nickerson DA, Eichler EE: Mutational and selective effects on copy-number variants in the human genome. *Nat Genet* 2007; **39**: S22–S29.
5 Perry GH, Tchinda J, McGrath SD *et al*: Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci USA* 2006; **103**: 8006–8011.
6 Sharp AJ, Locke DP, McGrath SD *et al*: Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005; **77**: 78–88.
7 Shaw CJ, Bi W, Lupski JR: Genetic proof of unequal meiotic crossovers in reciprocal deletion and duplication of 17p11.2. *Am J Hum Genet* 2002; **71**: 1072–1081.
8 Emanuel BS, Saitta SC: From microscopes to microarrays: dissecting recurrent chromosomal rearrangements. *Nat Rev Genet* 2007; **8**: 869–883.
9 Carter NP: Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* 2007; **39**: S16–S21.
10 Fan YS, Jayakar P, Zhu H *et al*: Detection of pathogenic gene copy number variations in patients with mental retardation by genomewide oligonucleotide array comparative genomic hybridization. *Hum Mutat* 2007; **28**: 1124–1132.
11 Redon R, Ishikawa S, Fitch KR *et al*: Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–454.
12 Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjold M: Detection of chromosomal imbalances in children with idiopathic mental retardation by array based comparative genomic hybridisation (array-CGH). *J Med Genet* 2005; **42**: 699–705.
13 Sharp AJ, Cheng Z, Eichler EE: Structural variation of the human genome. *Annu Rev Genomics Hum Genet* 2006; **7**: 407–442.
14 Trask BJ, Friedman C, Martin-Gallardo A *et al*: Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum Mol Genet* 1998; **7**: 13–26.
15 van Karnebeek CD, Scheper FY, Abeling NG *et al*: Etiology of mental retardation in children referred to a tertiary care center: a prospective study. *Am J Ment Retard* 2005; **110**: 253–267.
16 Rauch A, Hoyer J, Guth S *et al*: Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A* 2006; **140**: 2063–2074.
17 Zahir F, Friedman JM: The impact of array genomic hybridization on mental retardation research: a review of current technologies and their clinical utility. *Clin Genet* 2007; **72**: 271–287.
18 Shaikh TH: Oligonucleotide arrays for high-resolution analysis of copy number alteration in mental retardation/multiple congenital anomalies. *Genet Med* 2007; **9**: 617–625.
19 Rauch A, Ruschendorf F, Huang J *et al*: Molecular karyotyping using an SNP array for genomewide genotyping. *J Med Genet* 2004; **41**: 916–922.
20 Greshock J, Feng B, Nogueira C *et al*: A comparison of DNA copy number profiling platforms. *Cancer Res* 2007; **67**: 10173–10180.
21 McCarroll SA, Kuruvilla FG, Korn JM *et al*: Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* 2008; **40**: 1166–1174.
22 Hehir-Kwa JY, Egmont-Petersen M, Janssen IM, Smeets D, van Kessel AG, Veltman JA: Genome-wide copy number profiling on high-density bacterial artificial chromo-somes, single-nucleotide polymorphisms, and oligonucleotide microarrays: a platform comparison based on statistical power analysis. *DNA Res* 2007; **14**: 1–11.
23 Wong KK, deLeeuw RJ, Dosanjh NS *et al*: A comprehensive analysis of common copy-number variations in the human genome. *Am J Hum Genet* 2007; **80**: 91–104.
24 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.
25 Korn JM, Kuruvilla FG, McCarroll SA *et al*: Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nat Genet* 2008; **40**: 1253–1260.
26 Bernardini L, Castori M, Capalbo A *et al*: Syndromic craniosynostosis due to complex chromosome 5 rearrangement and MSX2 gene triplication. *Am J Med Genet A* 2007; **143A**: 2937–2943.
27 Carbone A, Bernardini L, Valenzano F *et al*: Array-based comparative genomic hybridization in early-stage mycosis fungoides: recurrent deletion of tumor suppressor genes BCL7A, SMAC/DIABLO, and RHOF. *Genes Chromosomes Cancer* 2008; **47**: 1067–1075.
28 Wagenstaller J, Spranger S, Lorenz-Depiereux B *et al*: Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. *Am J Hum Genet* 2007; **81**: 768–779.
29 Coe BP, Ylstra B, Carvalho B, Meijer GA, Macaulay C, Lam WL: Resolving the resolution of array CGH. *Genomics* 2007; **89**: 647–653.
30 Vermeesch JR, Fiegler H, de Leeuw N *et al*: Guidelines for molecular karyotyping in constitutional genetic diagnosis. *Eur J Hum Genet* 2007; **15**: 1105–1114.
31 Lee C, Iafrate AJ, Brothman AR: Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. *Nat Genet* 2007; **39**: S48–S54.
32 Hernandez-Martin A, Gonzalez-Sarmiento R, De Unamuno P: X-linked ichthyosis: an update. *Br J Dermatol* 1999; **141**: 617–627.
33 Shaw-Smith C, Redon R, Rickman L *et al*: Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and dupli-cations in patients with learning disability/mental retardation and dysmorphic fea-tures. *J Med Genet* 2004; **41**: 241–248.
34 Battaglia A, Hoyme HE, Dallapiccola B *et al*: Further delineation of deletion 1p36 syndrome in 60 patients: a recognizable phenotype and common cause of develop-mental delay and mental retardation. *Pediatrics* 2008; **121**: 404–410.
35 Heilstedt HA, Ballif BC, Howard LA *et al*: Physical map of 1p36, placement of breakpoints in monosomy 1p36, and clinical characterization of the syndrome. *Am J Hum Genet* 2003; **72**: 1200–1212.
36 Ballif BC, Yu W, Shaw CA, Kashork CD, Shaffer LG: Monosomy 1p36 breakpoint junctions suggest pre-meiotic breakage-fusion-bridge cycles are involved in generat-ing terminal deletions. *Hum Mol Genet* 2003; **12**: 2153–2165.
37 Perry GH, Ben-Dor A, Tsalenko A *et al*: The fine-scale and complex architecture of human copy-number variation. *Am J Hum Genet* 2008; **82**: 685–695.
38 Froyen G, Van Esch H, Bauters M *et al*: Detection of genomic copy number changes in patients with idiopathic mental retardation by high-resolution X-array-CGH: important role for increased gene dosage of XLMR genes. *Hum Mutat* 2007; **28**: 1034–1042.
39 Wolff DJ, Schwartz S, Carrel L: Molecular determination of X inactivation pattern correlates with phenotype in women with a structurally abnormal X chromosome. *Genet Med* 2000; **2**: 136–141.
40 Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I *et al*: Recurrent CNVs disrupt three candidate genes in schizophrenia patients. *Am J Hum Genet* 2008; **83**: 504–510.
41 Sudo K, Ito H, Iwamoto I, Morishita R, Asano T, Nagata K: SEPT9 sequence alternations causing hereditary neuralgic amyotrophy are associated with altered interactions with SEPT4/SEPT11 and resistance to Rho/Rhotekin-signaling. *Hum Mutat* 2007; **28**: 1005–1013.
42 Xie Y, Vessey JP, Konecna A, Dahm R, Macchi P, Kiebler MA: The GTP-binding protein Septin 7 is critical for dendrite branching and dendritic-spine morphology. *Curr Biol* 2007; **17**: 1746–1751.
43 Buser AM, Erne B, Werner HB, Nave KA, Schaeren-Wiemers N: The septin cytoske-leton in myelinating glia. *Mol Cell Neurosci* 2009; **40**: 156–166.
44 Zhang ZF, Ruivenkamp C, Staaf J *et al*: Detection of submicroscopic constitutional chromosome aberrations in clinical diagnostics: a validation of the practical perfor-mance of different array platforms. *Eur J Hum Genet* 2008; **16**: 786–792.