

ARTICLE

Detection of submicroscopic constitutional chromosome aberrations in clinical diagnostics: a validation of the practical performance of different array platforms

Zhong-Fa Zhang¹, Claudia Ruivenkamp², Johan Staaf³, Hongbo Zhu⁴, Michela Barbaro⁵, David Petillo¹, Sok Kean Khoo¹, Åke Borg³, Yao-Shan Fan⁴ and Jacqueline Schoumans^{*,5}

¹Laboratory of Cancer Genetics, Van Andel Research Institute, Grand Rapids, MI, USA; ²Department of Human and Clinical Genetics, Leiden University Medical Center (LUMC), Leiden, The Netherlands; ³Department of Oncology, Lund University, Lund, Sweden; ⁴The Dr John T Macdonald Foundation Center for Medical Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; ⁵Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden

For several decades etiological diagnosis of patients with idiopathic mental retardation (MR) and multiple congenital anomalies (MCA) has relied on chromosome analysis by karyotyping. Conventional karyotyping allows a genome-wide detection of chromosomal abnormalities but has a limited resolution. Recently, array-based comparative genomic hybridization (array CGH) technologies have been developed to evaluate DNA copy-number alterations across the whole-genome at a much higher resolution. It has proven to be an effective tool for detection of submicroscopic chromosome abnormalities causing congenital disorders and has recently been adopted for clinical applications. Here, we investigated four high-density array platforms with a theoretical resolution ≤ 100 kb: 33K tiling path BAC array, 500K Affymetrix SNP array, 385K NimbleGen oligonucleotide array and 244K Agilent oligonucleotide array for their robustness and implementation in our diagnostic setting. We evaluated the practical performance based on the detection of 10 previously characterized abnormalities whose size ranged from 100 kb to 3 Mb. Furthermore, array data analysis was performed using four computer programs developed for each corresponding platform to test their effective ability of reliable copy-number detection and their user-friendliness. All tested platforms provided sensitive performances, but our experience showed that accurate and user-friendly computer programs are of crucial importance for reliable copy-number detection.

European Journal of Human Genetics (2008) 16, 786–792; doi:10.1038/ejhg.2008.14; published online 20 February 2008

Keywords: submicroscopic chromosome aberrations; array platform comparison; copy-number detection

Introduction

Cytogenetic analysis for etiological diagnosis of patients with developmental delay and birth defects has since the

1970s relied on chromosome banding by karyotyping for whole-genome analysis. However, conventional karyotyping allows genome-wide detection of chromosomal abnormalities at a limited resolution (5–10 Mb). This problem can be overcome by using high-resolution screening technologies such as the recently developed array-based comparative genomic hybridization (array CGH). This technique is based on the competitive hybridization of reference and patient DNA samples to an immobilized

*Correspondence: Dr J Schoumans, Department of Molecular Medicine and Surgery, Karolinska Institute, Karolinska University Hospital Solna, Stockholm, Sweden. Tel: +46 8 51770453; Fax: +46 8 51773620; E-mail: Jacqueline.schoumans@ki.se

Received 25 September 2007; revised 17 December 2007; accepted 11 January 2008; published online 20 February 2008

target sequence on a glass slide or other solid platform. It allows a simultaneous evaluation of DNA copy-number alterations associated with chromosome abnormalities across the whole genome. Over the past years, array CGH has proven to be a powerful tool for detection of submicroscopic chromosome abnormalities in patients with idiopathic mental retardation (MR) and/or multiple congenital anomalies (MCA).^{1–10} These studies have led to the transfer of this research tool into a diagnostic instrument and a rapidly increasing number of clinical genetic laboratories offering array CGH as a genetic diagnostic service. The early stage experiences using array CGH were mostly based on 'in-house' produced bacterial artificial chromosome (BAC) arrays consisting of large-insert clones with an initial coverage of approximately one clone per Mb.^{2–5,7,8} With the majority of causal submicroscopic alterations detected randomly distributed across the genome,^{1–10} it is clear that whole-genome microarray is preferred over targeted microarray when investigating idiopathic MR. The added value of genome-wide high-resolution array CGH analysis over targeted arrays was recently discussed by Veltman and de Vries.¹¹

Recent developments in genome-wide array CGH technologies using oligonucleotides and single-nucleotide polymorphisms (SNPs) have resulted in a new generation of genome-wide array platforms. Several commercially available platforms containing a larger number of shorter DNA fragments (oligonucleotides) can now be used to interrogate the whole genome at an increased resolution. However, an adequate description of the capability of each platform has been difficult to define since the resolution of the array is not only determined by the number and size of probes, but more importantly by the genomic spacing and the hybridization sensitivity of the probes on the array. This makes platform comparison confusing as the increased number of array elements does not automatically result in a linear increase of performance that gives an experimental resolution that deviates from the theoretical resolution. The aim of this study was to validate the performance of four high-density array platforms for copy-number detection, based on the ability to detect submicroscopic constitutional chromosome abnormalities with the purpose of implementing a commercially available platform into our clinical diagnostic setting.

Materials and methods

Material selection

To investigate the performance of four different array CGH platforms based on their detection of causal genomic imbalances, we selected cases with well-characterized submicroscopic constitutional chromosome aberrations. Eight cases consisting of 10 deletions and/or duplications ranging in size from 100 kb to 3.0 Mb with different genomic locations were used in this study (Table 1). The

majority of genomic imbalances (6 out of 10) were located on chromosomes 22 and 17 since these two chromosomes are known to harbor a large number of genomic structures, such as segmental duplications that predispose for genomic rearrangements. The 10 abnormalities consisted of 3 interstitial microduplications (sizing 500, 637 and 800 kb), 3 interstitial microdeletions (sizing 1.5, 1.6 and 3.0 Mb), 3 terminal deletions (sizing 100 kb, 1.7 and 2.3 Mb) and 1 terminal duplication (1.3 Mb). Chromosome imbalances of cases 1, 4, 5, 6 and 7 had initially been identified using 33K BAC array CGH platform, while the abnormalities of cases 2, 3 and 8 had been detected by subtelomeric fluorescence *in situ* hybridization (FISH). All abnormalities had been confirmed and further mapped to ~100 kb accuracy using locus-specific BAC FISH or synthetic probe multiplex ligation-dependent probe amplification. The amount of cells having a deletion in the mosaic case (case 5) had been determined using interphase FISH by investigation of 200 nuclei. Two of the rearrangements (637 kb duplication at Xp21.2 and 100 kb deletion at 22q13.3) were accurately mapped to the base pair level by PCR and sequencing,^{12,13} and five of the cases (cases 2, 3, 6, 7 and 8) have been previously reported.^{13–17} To investigate the performance according to the ability to detect small but clinically significant genomic imbalances, all selected abnormalities included in this study were most likely causal for the phenotypes in the patients except for the *de novo* 500-kb duplication on 9p in case 6 that had no reported normal variants in this region at the time of investigation and was reported with an unclear clinical significance.¹⁷ They were all *de novo* genomic alterations (except case 7, which had a causal maternally inherited duplication¹³), and none of the imbalances had previously been reported as a normal variant except the duplication in case 6.

DNA preparation

Genomic DNA was extracted from blood samples or Epstein–Barr virus-transformed lymphocytes using Puregene blood kit (Gentra Systems Inc., Minneapolis, MN, USA) according to the manufacturer's protocol or by classical phenol/chloroform protocol. The reference genomic DNA consisted of a pool of 10 normal male or 10 normal female subjects (Promega, Madison, WI, USA).

Tiling path BAC array

The 33K tiling path BAC array with complete genome coverage containing 33 370 large-insert clones produced by the Swegene DNA Microarray Resource Center, Department of Oncology, Lund University, Sweden (<http://swegene.onk.lu.se>) was used. The clone set consisted of the 32K BAC clone library (CHORI BACPAC Resources, see Web resources),¹⁸ with additional clones located in the telomeric regions¹⁹ and clones covering microdeletion syndromes.⁵ The arrays were printed as previously described.²⁰ BAC clones were mapped according to the UCSC Human

Table 1 Array results (all genomic position are mapped according to NCBI Build 35)

Case	Array platform/software used for analysis	Characterized chromosomal abnormalities	33K tiling BAC array/BASE v 1.2			500K Affymetrix array/CNAG v 2.0/dChip+R			385K NimbleGen array/SignalMap v 1.8			244K Agilent array/CGH Analytics		
			Start pos	Stop pos	Size (Mb)	Start pos	Stop pos	Size (Mb)	Start pos	Stop pos	Size (Mb)	Start pos	Stop pos	Size (Mb)
1	3.0 Mb del (17)(p13.2p13.1)		4460333	7459912	2.9	4518438	7409662	2.9	4511437	7387669	2.9	4510789	7395371	2.9
2	2.3 Mb del (12)(q24.33)		130039984	132389806	2.3	130112899	132388245	2.3	130112899	132389474	2.3	13011326	132389426	2.3
3	1.3 Mb dup (17)(q25.3)		77390101	78398677	0.9	77684980	78605724	1.3	77384569	78636387	1.3	77286555	78653589	1.3
3	1.7 Mb del (6)(p25.3)		64779	1756785	1.7	110141	1795693	1.8	58851	1804244	1.7	108083	1808508	1.7
4	1.5 Mb del (22)(q11.21q11.22)		19833746	21440190	1.5	19784312	21277043	1.5	19785663	21287033	1.5	19829912	21287554	1.5
5	1.6 Mb del (22)(q11.21q11.21) mosaic		17383746	19028374	1.3 ^a	17311817	18574383	ND	—	—	—	17080555	18667748	1.6
6	800 kb dup (22)(q11.21q11.21)		19022374	19849078	0.7	19057113	19784812	0.8	19054053	19800518	0.8	19033444	19829971	0.8
6	500 kb dup (9)(p24.3p24.3)		118663	660178	0.6	36181	679590	ND	—	—	—	194193	700834	0.5
7	637 kb ^b dup (X)(p21.2p21.2)		30014387	30681589	0.6	30000217	30609083	0.6	29995709	30636022	0.6	30003032	30613450	0.6
8	100 kb del (22)(q13.33)		—	—	Failed	—	—	—	—	—	—	49427470	49509094	0.1

ND, not detected; NT, not tested due to probe coverage.

^aNot detected in CNAG analysis but visible in dChip/R.^bStart-stop position determined by sequencing 30145030 – 30781736.

Genome browser May 2004. Sample labeling and hybridization were performed as described previously,²¹ and the arrays were scanned using GenePix® Professional 4200A scanner (Axon Instruments, Union City, CA, USA). Identification of individual spots on scanned arrays was performed with GenePix Pro 6.0 (Axon Instruments) and the quantified data matrix was loaded into BioArray Software Environment BASE.²² For breakpoint identification, a BASE plug-in GLAD²³ was used and the threshold for gains and losses was set to ≥ 3 consecutive clones with \log_2 (ratio) of ± 0.2 . Color reverse experiments were performed.

Affymetrix 500K GeneChip

The Affymetrix 500K GeneChip array contains 25-mer oligonucleotides distributed over two subarrays containing 262 264 and 238 304 SNPs, respectively, representing a total of 500 568 SNPs was used. The median SNP spacing is ~ 2.5 kb. Each subarray interrogates SNPs residing *NspI* or *StyI* PCR amplicons that range in size from 200 to 1000 bp. Experiments were performed according to protocols provided by the manufacturer (Affymetrix Inc., Santa Clara, CA, USA). Raw copy numbers were estimated using CNAG software version 2.0 available at <http://www.genome.u-min.jp/CNAGtop2.html>.²⁴ SNP data of 48 samples were downloaded from Affymetrix (http://www.affymetrix.com/support/technical/sample_data/500K_data.affx) and were used as unpaired normal references. In addition, the raw copy-number values were calculated using the dChip software available at <http://biosun1.harvard.edu/complab/dchip/>,²⁵ and then a homemade script in R was applied to plot the values per chromosome across the genomic location for a more clear visualization of the imbalances. Each 250K array was analyzed separately, but for a more accurate breakpoint determination of each copy-number alteration, results from both chips were combined.

NimbleGen 385K oligonucleotide array

The 385K oligonucleotide array produced by NimbleGen Systems Inc. (Madison, WI, USA) was used. The array contains 386 165 isothermal oligonucleotides probes (45- to 85-mer with a median probe spacing of 6 kb) with complete genome coverage. Probe design, array fabrication, array CGH experiments including DNA labeling, hybridization, array scanning, data normalization and \log_2 copy-number ratio calculation were performed by NimbleGen Systems Inc. Array data were analyzed using the SignalMap Software version 1.8 (NimbleGen Systems Inc.).

Agilent 244K oligonucleotide array

The 244K oligonucleotide array with complete genome coverage produced by Agilent Technologies (Wilmington, DE, USA) was used. The array contains 236 000 oligonucleotides probes (60-mer) plus 1000 triplicates and 5000

controls with a median probe spacing of 8.9 kb. Experiments were performed according to the manufacturer's protocol. After hybridization and washing, the slides were scanned on an Agilent Microarray Scanner. Captured images were analyzed with Feature Extraction Software v 9.1 and CGH Analytics 3.4 (Agilent Technologies) as described previously by Fan *et al.*⁹

Array performance

All hybridizations using the different platforms were successful and passed the quality criteria in each software package except for one hybridization on the Affymetrix 500K array (case 8), which most likely failed due to too low quantity of DNA (<500 ng). One microgram of genomic DNA was used for each hybridization (except for the Affymetrix platform, where 250 ng was used for each enzyme assay). Color reverse experiments were performed on the 33K tiling BAC array using 1 μ g for each labeling and single hybridizations were performed on the other platforms. Labeling, hybridization, washing and scanning of the slides of each array platform were performed according to the recommendation of each manufacturer. Array CGH analysis was carried out using four different computer programs, developed and/or recommended for each platform. For two of the platforms, software especially developed for array CGH application were commercially available (CGH Analytics from Agilent and SignalMap from Nimblegen), while for the two other platforms we relied on academically developed software available online and free for public download. All abnormalities were analyzed blindly and the results of the array experiments are summarized in Table 1.

Results

33K tiling path BAC arrays

The 33K tiling path array detected all abnormalities that were hybridized (cases 1–7). Case 8, with a subtelomeric deletion of 100 kb, was not tested due to limited DNA quantity, and since the array only contained one clone in the 22q13.33 deleted region it was not expected to be detectable according to the threshold used in our analysis (three consecutive clones exceeding threshold \log_2 ratio ± 0.2). Using the 33K BAC array, we observed an average CNV (copy-number variant present in healthy individuals) of eight per individual. In all, 6 out of 10 abnormalities were initially identified by the 33K BAC; therefore, this platform could not objectively be compared with the other three platforms.

Affymetrix 500K GeneChip

One out of eight 500K Affymetrix SNP array experiments failed to give good quality data (case 8 containing a 100-kb 22q13.33 deletion). Unfortunately, the experiment could not be repeated due to insufficient DNA quantity.

Therefore, it remains unclear whether the abnormality would have been detected due to the low coverage of probes (two probes on *NspI* and three probes on *StyI*) in the region on the 500K SNP array. First, copy-number analysis was blindly carried out using CNAG v 2.0. One abnormality, the 1.6-Mb mosaic deletion at 22q11.21, failed to be detected in this analysis (case 5, see Table 1). Second, analysis was performed using a dChip and an R script for a clearer visualization of the genomic imbalances for each chromosome. By visual inspection of the genomic region containing the expected abnormality, the 1.6-Mb deletion missed by the CNAG analysis was visible (Figure 1). An average of 20 CNVs per individual was observed using this platform.

NimbleGen 385K oligonucleotide array

All experiments on NimbleGen arrays were performed at NimbleGen Systems Iceland (Reykjavik, Iceland), and data were analyzed using SignalMap version 1.8 software. Eight hybridizations were performed using four arrays. Samples 3, 5, 6 and 8 were hybridized on 'reuse' arrays that were previously used on samples 1, 2, 4 and 7, respectively. The arrays were cleaned from genomic DNA prior to the second hybridization. Three abnormalities escaped detection using this platform; a 100-kb 22q13.33 deletion, a 500-kb 9p24.3 duplication and a 1.6-Mb mosaic 22q11.21, while all three regions had a dense probe coverage on the array. However, the three samples containing these abnormalities were all hybridized on a 'reuse' array. An average of 20 CNVs per individual was observed in the tested samples.

Agilent 244K oligonucleotide array

All abnormalities were correctly detected using the 244K Agilent platform and the CGH Analytics software. An average of 19 CNVs per individual was observed.

Discussion

Genome-wide copy-number detection using microarrays is becoming an indispensable genetic analysis in the diagnosis of idiopathic MR/MCA. Guidelines for molecular karyotyping in constitutional genetic diagnosis have been published.²⁶ However, the practical implementation of molecular karyotyping into the cytogenetic laboratory, for at least partial replacement of conventional karyotyping, is not an easy road. Bioinformaticians, to assist in array CGH analysis by designing specialized algorithms for reliable detection of copy-number alteration, are usually not accessible in a routine cytogenetic laboratory. Therefore, it is not only of great importance that arrays are subjected to test their reproducibility, but also the software used for analysis should be user-friendly and reliable for detection of gains and losses in the genome.

Four different array platforms for copy-number detection were tested to investigate which of the high-density array

platforms would be most suitable for implementation in our diagnostic setting. Microarray platform comparison is complex since its resolution is not only determined by the number, size and spacing of the array elements, but also depends on the signal-to-noise ratio of each probe. Statistical calculation to estimate the functional resolution instead of the theoretical resolution of various platforms have recently been reported.^{27,28} We validated the practical performance of different CGH platforms based on the detection of submicroscopic chromosome imbalances identified in patients with MR and/or birth defects, since we are targeting this group of patients by using high-resolution genome-wide array analysis in our clinic.

Using the 33K tiling path BAC array, we detected on average 8 CNVs, while the other three platforms with increased resolution detected 19–20 CNVs per individual. A drawback of genome-wide analysis at increased resolution is the increased detection of inherited submicroscopic CNVs from phenotypically normal parents, reflecting normal CNVs rather than disease-associated genomic changes. This initially complicated the discernment between a copy-number alteration that causes disease *versus* one without clinical consequences. However, since the first reports on normal large-scale copy-number variation in 2004,^{29,30} knowledge has dramatically improved by cataloging normal variation in several ethnic populations.^{31,32} By making the tremendous number of detected benign CNVs publicly accessible in a database at <http://projects.tcag.ca/variation/>, array analysis has been greatly facilitated. All CNVs observed in these eight patients were listed in the database of genomic variants and were therefore not subjected to further analysis.

Using the 33K BAC array, nine out of nine abnormalities were detected (case 8 was not tested). However, six of the abnormalities were initially identified by this platform, and therefore the comparison to the other platforms was biased. In addition, the data from the 33K BAC were obtained from replicate analyses, while all other array data were obtained from a single hybridization per chip. The replicate analysis increases the reliability of results from the BAC array, while a single hybridization would be more sensitive for false positive results. Using the Agilent 244K array, 10 out of 10 abnormalities were correctly identified. The 385K NimbleGen array detected 7 out of 10. The 500K

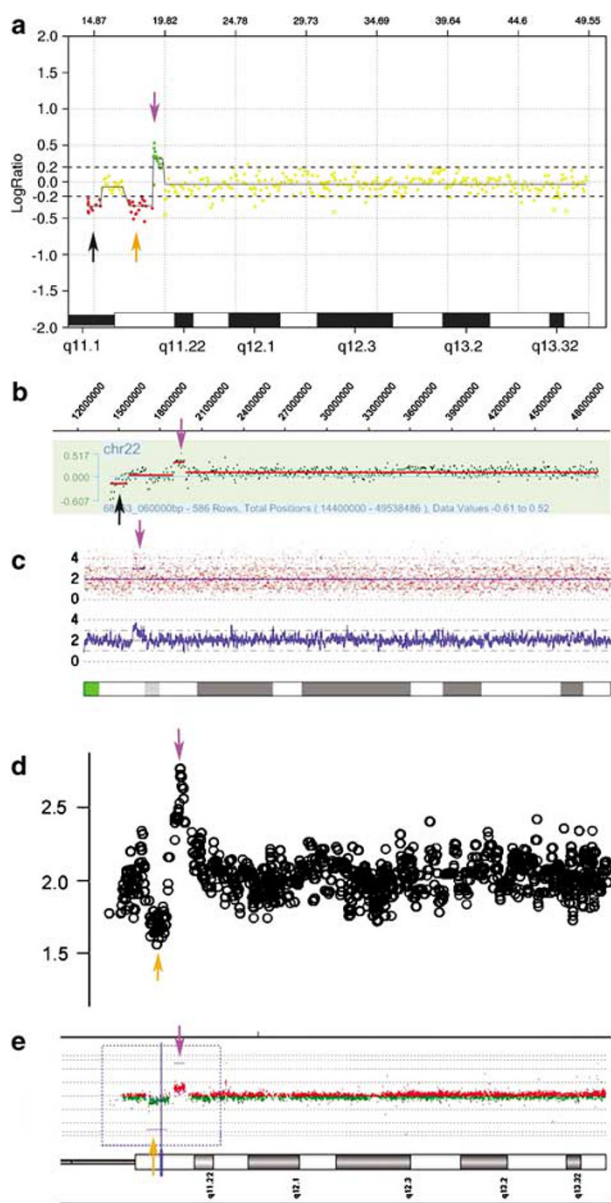


Figure 1 Displays of array CGH plots from the four different platforms. Chromosome 22 from case 5 is shown, which contains a 1.6-Mb mosaic deletion (70% of the cells) with genomic location shown in 17.3–19.0 Mb (marked by an orange arrow) and a 800-kb duplication from position 19.0–19.8 Mb (marked by a purple arrow). In addition, the patient has a common CNV at the centromere position (black arrow). (a) Represents the plot from the analysis by the 33K tiling path BAC array performed in BASE using breakpoint identification plug-in 'GLAD.' The results show the detection of the CNV at the centromere, the 1.6-Mb deletion (clones displayed in red) and the 800-kb duplication (clones displayed in green). The 1.6-Mb deletion shows a mean ratio value of 0.3, which indicates a mosaic deletion. (b) Displays a CGH plot from the 385K NimbleGen array analyzed by SignalMap. The CNV at the centromeric region and the duplication are detected (visualized by red bar deviating from normal). The mosaic deletion is not observed. (c) The plot of Affymetrix 500K data, using the CNAG analysis software (*NspI* assay) is shown. The CNV region is excluded from analysis and the 800-kb duplication is visible by the blue bar deviating from normal. The 1.6-Mb deletion is not detected. (d) The Affymetrix *NspI* SNP array data after analysis in dChip and R. Abnormalities are detected by visual inspection. The 800-kb duplication is clearly displayed, while a 1.6-Mb deletion is less clearly visible. (e) In this image, the plot of the Agilent 244K is shown using CGH Analytics software. The 1.6-Mb deletion and the 800-kb duplication are both observed (visible by the blue bar deviating from normal) and the plot indicates a mosaic deletion. The CNV at the centromeric region is not detected since only a few probes on the array were located in that region.

SNP Affymetrix array detected eight out of nine using the CNAG software and nine out of nine using dChip in combination with an R script. The failure to detect the mosaic deletion by the two platforms mentioned above is most likely due to the fact that the Affymetrix and NimbleGen analyses displayed a slightly noisier data (Figure 1), compared with the BAC and the Agilent analyses. Despite the relatively large size of the deletion (1.6 Mb), it was only present in 70% of the cells. This might indicate that mosaic cases of small chromosome segments might slightly easier escape detection using the Affymetrix and the NimbleGen platform and their corresponding software. However, the detection of copy-number imbalances in samples containing different cell populations can most likely be improved for the Affymetrix platform by the use of in-house control references instead of the reference data from Affymetrix. Thus, the choice of reference samples significantly affects the copy-number ratios. It is therefore strongly recommended to use reference samples that represent the experiment conditions of the tested samples to increase the sensitivity of the array. On the other hand, the need for optimization of the analysis by using a large series of in-house performed controls makes the platform less user-friendly for implementation in a clinical setting.

The 500-kb 9p24.3 duplication was not detected by the NimbleGen array. This can be explained by the fact that the theoretical ratio of a duplication is three copies from the patient *versus* two copies from the reference, which is closer to the random noise level compared with deletions (ratio one copy from the patient *versus* two copies from the reference). Very small duplications are thus more difficult to discriminate from experimental noise. Finally, the 100-kb deletion at 22q13.33 was not detected by either the NimbleGen or Affymetrix array, most probably due to the quality of DNA.

Regarding computerized array data analysis using the different software packages, the commercially developed software from Agilent Technologies (CGH Analytics) and NimbleGen Inc. (SignalMap) were considered to be more user-friendly compared with the free downloadable software packages. The CGH Analytics software and BASE produced good illustrations of the abnormalities, which greatly facilitated the detection of copy-number changes (see Figure 1). As expected, the size of each aberration was found to be slightly variable using different platforms (Table 1), because it is determined by the genomic position of the array elements of each platform, which have different coverage and distribution.

An important advantage to consider for the Affymetrix SNP array over the other three platforms is the combination of loss-of-heterozygosity analysis together with the CGH analysis, which enables the detection of copy-number neutral chromosomal aberrations such as uniparental disomy.

In conclusion, the four platforms we tested provided good and sensitive performances. However, we observed a variable capacity between the different platforms to detect the submicroscopic genomic alterations based on the different software analysis used. For the transition of the array technology into the diagnostic laboratories, and the partial replacement of conventional karyotyping by molecular karyotyping, the laboratory personnel needs to be retrained to be able to perform array analysis on an increasing scale. Therefore, reliable and user-friendly computer programs are of crucial importance. In our study, we found the array software package from Agilent Technologies to be the most accurate and user-friendly.

Acknowledgements

The work was financially supported by the Swedish Research Council, Frimurarna Barnhuset Foundation and the Gerber Foundation.

References

- 1 de Vries BB, Pfundt R, Leisink M *et al*: Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 2005; **77**: 606–616.
- 2 Schoumans J, Ruivenkamp C, Holmberg E, Kyllerman M, Anderlid BM, Nordenskjöld 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.
- 3 Shaw-Smith C, Redon R, Rickman L *et al*: Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet* 2004; **41**: 241–248.
- 4 Tyson C, Harvard C, Locker R *et al*: Submicroscopic deletions and duplications in individuals with intellectual disability detected by array-CGH. *Am J Med Genet A* 2005; **139**: 173–185.
- 5 Vissers LE, de Vries BB, Osoegawa K *et al*: Array-based comparative genomic hybridization for the genomewide detection of submicroscopic chromosomal abnormalities. *Am J Hum Genet* 2003; **73**: 1261–1270.
- 6 Friedman JM, Baross A, Delaney AD *et al*: Oligonucleotide microarray analysis of genomic imbalance in children with mental retardation. *Am J Hum Genet* 2006; **79**: 500–513.
- 7 Menten B, Maas N, Thienpont B *et al*: Emerging patterns of cryptic chromosomal imbalance in patients with idiopathic mental retardation and multiple congenital anomalies: a new series of 140 patients and review of published reports. *J Med Genet* 2006; **43**: 625–633.
- 8 Rosenberg C, Knijnenburg J, Bakker E *et al*: Array-CGH detection of micro rearrangements in mentally retarded individuals: clinical significance of imbalances present both in affected children and normal parents. *J Med Genet* 2006; **43**: 180–186.
- 9 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.
- 10 Slater HR, Bailey DK, Ren H *et al*: High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs. *Am J Hum Genet* 2005; **77**: 709–726.
- 11 Veltman JA, de Vries BB: Whole-genome array comparative genome hybridization: the preferred diagnostic choice in post-natal clinical cytogenetics. *J Mol Diagn* 2007; **9**: 277.
- 12 Bonaglia MC, Giorda R, Mani E *et al*: Identification of a recurrent breakpoint within the SHANK3 gene in the 22q13.3 deletion syndrome. *J Med Genet* 2006; **43**: 822–828.

- 13 Barbaro M, Oscarson M, Schoumans J, Staaf J, Ivarsson SA, Wedell A: Isolated 46,XY gonadal dysgenesis in two sisters caused by a Xp21.2 interstitial duplication containing the DAX1 gene. *J Clin Endocrinol Metab* 2007; **92**: 3305–3313.
- 14 Anderlid BM, Schoumans J, Anneren G *et al*: FISH-mapping of a 100-kb terminal 22q13 deletion. *Hum Genet* 2002; **110**: 439–443.
- 15 Anderlid BM, Schoumans J, Hallqvist A *et al*: Cryptic subtelomeric 6p deletion in a girl with congenital malformations and severe language impairment. *Eur J Hum Genet* 2003; **11**: 89–92.
- 16 Schoumans J, Sanner G, Nordenskjold M, Anderlid BM: Detailed clinical description of four patients with 1.3 and 2.1 Mb chromosome imbalances derived from a familial t(12;17)(q24.33;q25.3). *Am J Med Genet A* 2005; **134**: 254–258.
- 17 Schoumans J, Wincent J, Barbaro M *et al*: Comprehensive mutational analysis of a cohort of Swedish Cornelia de Lange syndrome patients. *Eur J Hum Genet* 2006; **15**: 143–149.
- 18 Krzywinski M, Bosdet I, Smailus D *et al*: A set of BAC clones spanning the human genome. *Nucleic Acids Res* 2004; **32**: 3651–3660.
- 19 Knight SJ, Lese CM, Precht KS *et al*: An optimized set of human telomere clones for studying telomere integrity and architecture. *Am J Hum Genet* 2000; **67**: 320–332.
- 20 Jonsson G, Staaf J, Olsson E *et al*: High-resolution genomic profiles of breast cancer cell lines assessed by tiling BAC array comparative genomic hybridization. *Genes Chromosomes Cancer* 2007; **46**: 543–558.
- 21 Schoumans J, Johansson B, Corcoran M *et al*: Characterisation of dic(9;20)(p11–13;q11) in childhood B-cell precursor acute lymphoblastic leukaemia by tiling resolution array-based comparative genomic hybridisation reveals clustered breakpoints at 9p13.2 and 20q11.2. *Br J Haematol* 2006; **135**: 492–499.
- 22 Saal LH, Troein C, Vallon-Christersson J, Gruvberger S, Borg A, Peterson C: BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. *Genome Biol* 2002; **3**: SOFTWARE0003.
- 23 Hupe P, Stransky N, Thiery JP, Radvanyi F, Barillot E: Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* 2004; **20**: 3413–3422.
- 24 Nannya Y, Sanada M, Nakazaki K *et al*: A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 2005; **65**: 6071–6079.
- 25 Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C: dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 2004; **20**: 1233–1240.
- 26 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.
- 27 Coe BP, Ylstra B, Carvalho B, Meijer GA, Macaulay C, Lam WL: Resolving the resolution of array CGH. *Genomics* 2007; **89**: 647–653.
- 28 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 chromosomes, single-nucleotide polymorphisms, and oligonucleotide microarrays: a platform comparison based on statistical power analysis. *DNA Res* 2007; **14**: 1–11.
- 29 Iafrate AJ, Feuk L, Rivera MN *et al*: Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949–951.
- 30 Sebat J, Lakshmi B, Troge J *et al*: Large-scale copy number polymorphism in the human genome. *Science* 2004; **305**: 525–528.
- 31 Redon R, Ishikawa S, Fitch KR *et al*: Global variation in copy number in the human genome. *Nature* 2006; **444**: 444–454.
- 32 Qiao Y, Liu X, Harvard C *et al*: Large-scale copy number variants (CNVs): distribution in normal subjects and FISH/real-time qPCR analysis. *BMC Genomics* 2007; **8**: 167.