Nucleic Acids Research, 2005, Vol. 33, No. 22 ***e192***
doi:10.1093/nar/gni191

Human and mouse oligonucleotide-based array CGH

Paul van den IJssel, Marianne Tijssen, Suet-Feung Chin[1], Paul Eijk, Beatriz Carvalho, Erik Hopmans, Henne Holstege[2], Dhinoth Kumar Bangarusamy[3], Jos Jonkers[2], Gerrit A. Meijer, Carlos Caldas[1] and Bauke Ylstra*

Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands, [1]Cancer Genomics Program, Department of Oncology, University of Cambridge, Cambridge, UK, [2]Division of Molecular Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands and [3]Genome Institute of Singapore, Biomedical Sciences Institutes, Singapore

Received July 19, 2005; Revised September 26, 2005; Accepted November 24, 2005

ABSTRACT

Array-based comparative genomic hybridization is a high resolution method for measuring chromosomal copy number changes. Here we present a validated protocol using in-house spotted oligonucleotide libraries for array comparative genomic hybridization (CGH). This oligo array CGH platform yields reproducible results and is capable of detecting single copy gains, multi-copy amplifications as well as homozygous and heterozygous deletions as small as 100 kb with high resolution. A human oligonucleotide library was printed on amine binding slides. Arrays were hybridized using a hybstation and analysed using BlueFuse feature extraction software, with >95% of spots passing quality control. The protocol allows as little as 300 ng of input DNA and a 90% reduction of Cot-1 DNA without compromising quality. High quality results have also been obtained with DNA from archival tissue. Finally, in addition to human oligo arrays, we have applied the protocol successfully to mouse oligo arrays. We believe that this oligo-based platform using 'off-the-shelf' oligo libraries provides an easy accessible alternative to BAC arrays for CGH, which is cost-effective, available at high resolution and easily implemented for any sequenced organism without compromising the quality of the results.

INTRODUCTION

Microarray comparative genomic hybridization (CGH) is now the method of choice for the detection of DNA copy number changes, such as gains, losses, amplifications and (micro-) deletions in tumours and genetic diseases (1,2). The first whole genome microarray contained 2400 large-insert genomic clones, primarily bacterial artificial chromosomes (BACs) (3). With the total human genome covering ~3000 Mb, the resolution of this array is on average close to 1 Mb, about one order of magnitude higher than that obtained with classical CGH (4). For a full coverage resolution array ~30 000 BACs have been arrayed (5), increasing the resolution with another order of magnitude. However, producing such large numbers of BACs for array CGH is expensive and time-consuming and due to the large size of the BACs, the limits of BAC array CGH resolution have been reached. These problems can be overcome when oligonucleotides are used as targets in microarray experiments. Oligonucleotides allow a sheer infinite resolution, great flexibility and are cost-effective (6). Moreover, oligonucleotides allow for the generation of microarrays for any organism of which the genome has been sequenced. Attempts have been undertaken to increase the resolution of BAC arrays in other ways, such as BAC-based (7) or exon-based (8) PCR-fragments, but both are laborious ways to generate targets for whole genome array CGH that cannot compete with the flexibility and versatility of oligonucleotides. Finally, oligonucleotide arrays are being used, designed and accepted for expression profiling and thus widely available. Using the same oligo array for CGH and expression profiling allows direct comparison of mRNA expression and DNA copy number ratios (9).

We have published results showing proof of principle for the use of spotted 60mer oligonucleotides for high resolution microarray CGH (6). Commercially available oligonucleotide microarray platforms have also successfully been used for array CGH (10–12). The aim of the current study was to optimize and validate our protocol for in-house spotted oligo array CGH. We present results showing high resolution detection of single copy gains, amplifications and (micro-) deletions. Fluorescence *in situ* hybridization (FISH) has

*To whom correspondence should be addressed at Microarray Core Facility, VU University Medical Center, van der Boechorststraat 7-9, 1081 BT Amsterdam, The Netherlands. Tel: +31 20 444 8299; Fax: +31 20 444 8318; Email: b.ylstra@vumc.nl

been used to validate both homozygous and heterozygous deletions. The presented oligo array CGH protocol provides a highly sensitive and reproducible platform both for human and mouse samples applicable to DNA isolated from both fresh and formalin fixed paraffin embedded (FFPE) tissue as validated by BAC arrays. For laboratories that already have oligonucleotide-based expression array platforms running or are planning to, we here provide protocols with which those platforms can easily be adapted for array CGH.

MATERIALS AND METHODS

Oligo array preparations

The Human Release 2.0 oligonucleotide library, containing 60mer oligonucleotides representing 28 830 unique genes and the Mouse Release 2.0 oligonucleotide library, containing 65mer oligonucleotides representing 21 587 unique genes as designed by Compugen (San Jose, CA, USA), were obtained from Sigma-Genosys (Zwijndrecht, The Netherlands). The oligonucleotides were dissolved at 10 μM concentration in 50 mM sodium phosphate buffer pH 8.5 and single spotted onto CodeLink™ slides (Amersham Biosciences), using an OmniGrid® 100 microarrayer (Genomic Solutions, Ann Arbor, MI, USA) equipped with SMP3 pins (TeleChem International, Sunnyvale, CA, USA). After printing slides were processed and blocked prior to use according to the manufacturer's protocol.

DNA isolation and labelling

Genomic DNA was isolated from blood obtained from 10 males or 10 females using DNAzol (Invitrogen) according to the manufacturer's protocol and was pooled for use as reference DNA. Genomic DNA from cell lines BT474 (ATCC nr. HTB-20), MDA-MB-468 (ATCC nr. HTB-132), SKBR7 (obtained from Dr M. J. O'Hare, Ludwig Institute for Cancer Research, London, UK), SUM159, GM01750, GM13031 and GM07408 (obtained from Coriell Institute for Medical Research, Camden, USA) was isolated according to Snijders *et al*. (3). Genomic DNA from FFPE gastric tumour was isolated according to Weiss *et al*. (13).

DNA labelling was performed essentially according to Snijders *et al*. (3) using the BioPrime DNA labelling system (Invitrogen). Genomic DNA (300 ng) was mixed with 20 μl of 2.5× Random primer (Invitrogen) and MilliQ water to a total volume of 42 μl. The mixture was then incubated for 10 min at 100°C to denature the DNA, instantly cooled down on ice/water and kept on ice while adding 5 μl of dNTP mixture (2 mM dATP, 0.5 mM dCTP, 2 mM dGTP, 2 mM dTTP, 10 mM Tris pH 7.6, 1 mM EDTA) 2 μl of 1 mM Cy3-dCTP (test sample) or Cy5-dCTP (pooled reference sample) (Perkin Elmer) and 1 μl of Klenow DNA polymerase (40 U/μl, Invitrogen). After gently mixing the DNA was incubated for 16 h at 37°C. Unincorporated label was removed before mixing test and reference samples using ProbeQuant G-50 Micro Columns (Amersham Biosciences) according to the manufacturer's protocol, by applying the sample to the top centre of the resin while being careful not to disturb the resin bed. The purified sample was stored in the dark until used on the same day or at −20°C for longer storage.

Hybridization

For preparation of the hybridization mixture of both human and mouse samples 50 μl of Cy3-labelled test DNA, 50 μl of Cy5-labelled reference DNA and 10 μg Human Cot-1 DNA (Invitrogen) were mixed and precipitated using 0.1 vol of 3 M NaAc pH 5.2 and 2.5 vol of ice-cold absolute ethanol. After mixing by inversion the DNA was collected by centrifugation for 30 min at 20 000 *g* and 4°C, the supernatant aspirated and the pellet air-dried for ∼5–10 min. The pellet was then dissolved in 13 μl Yeast tRNA (100 μg/μl, Invitrogen) and 26 μl 20% SDS taking care to prevent foam formation. After incubating at room temperature for 15 min 91 μl of Master mix [14.3% (w/v) dextran sulphate (USB), 50% (v/v) formamide (Invitrogen), 2.9× SSC pH 7.0 (Sigma)] was added and gently mixed. The hybridization solution was then incubated at 73°C for 10 min to denature the DNA and subsequently at 37°C for 60 min to allow the Cot-1 DNA to block repetitive sequences. Hybridization and washing were done automatically using a GeneTAC/HybArray12 hybstation (Genomic Solutions/Perkin Elmer). Hybridization was done for 38 h at 37°C. Subsequently slides were washed 6 cycles (flow for 10 s, hold for 20 s) with 50% (v/v) formamide, 2× SSC, 2 cycles with phosphate-buffer [0.1 M Na_2HPO_4/NaH_2PO_4, pH 8.0, 0.1% (v/v) Igepal CA630 (Sigma)], 2 cycles with 0.2× SSC (Sigma) and 2 cycles with 0.1× SSC. Slides were then taken out of the hybstation and briefly rinsed in 0.01× SSC, dried by centrifugation for 3 min at 1000 *g* and scanned using an Microarray Scanner G2505B (Agilent Technologies).

Analysis

Spot analysis and quality control were fully automated using BlueFuse version 3.1 (BlueGnome, Cambridge, UK). Spots were excluded when the quality flag was <1 or the Confidence value <0.1. Oligonucleotides from the human library were mapped to the human genome build NCBI35 and from the mouse library to the mouse genome build NCBIm33 (Ensembl v31). Oligonucleotide sequences and mapping have been made accessible by Compugen (San Jose, CA, USA) and The Sanger Institute (Hinxton, Cambridge, UK), respectively, via www.ensembl.org. A unique chromosomal position was identified for 26 845 of the 28 830 oligonucleotides in the human library. For the mouse library, 15 140 oligonucleotides of the 21 587 had a unique position. Oligonucleotides were excluded when they mapped to more than one position in the genome or showed one or more mismatches with regard to the current build.

Log2ratios of spots that were not excluded after quality flagging and mapping were normalized to their mode value. Weighted moving average values were then calculated using a triangular function and a window of 250 kb as described (11) solely to display genomic profiles where indicated. Breakpoints, gains, losses and amplifications were detected using the January 2005 version of ArrayCGHsmooth (14) that uses a ratio-dependent penalty for breakpoints in order to facilitate the detection of amplifications. Genetic Algorithm Parameters used were lambda: 3, pool size: 300, maximum number of breakpoints initial pool: 10 and maximum number of generations: 5000. The minimum difference between levels for the 'Join Closest Levels' parameter was 0.1. The 'new ratio join' levels were used to display in the figures.

Fluorescence *in situ* hybridization

FISH was performed using BAC clones identified for the deleted region of interest i.e. the *RB1* gene was identified from Ensembl to be contained within the RP11-305D15 BAC clone while the deleted region of 12q24 is represented by three BACs RP11-340F14, RP11-44F24 and RP11-7M8. The BAC DNA was isolated by the alkaline-lysis method and labelled with biotin-16-dUTP (Roche Biochemicals, UK) using nick translation (Vysis, UK). The labelled BAC was precipitated together with a whole chromosome 13 paint labelled with Spectrum Orange (Abbott Diagnostics, UK) in the presence of unlabelled Human Cot-1 DNA. The precipitated DNA was resuspended in hybridization buffer (50% formamide/10% dextran sulphate/2× SSC).

Metaphase preparations from MDA-MB-468 were obtained using standard potassium chloride/Carnoy's fixative protocols and stored at −20°C or from peripheral blood lymphocytes of normal individuals for reference. Slides were denatured in 70% formamide/2× SSC at 72°C for 1 min and immersed into ice-cold ethanol series (70, 90 and 100%) for 3 min each. A total of 15 µl of probes were denatured at 72°C for 10 min, allowed to pre-anneal at 37°C for 30 min and applied onto slides for a 72 h hybridization. Bound probes were detected with streptavidin-FITC (Vector Labs, UK) and counterstained with DAPI in a mounting medium (Vectashield, Vector Labs, UK). Slides were examined under a Zeiss Axioplan 2 epifluorescence microscope equipped with single band pass emission filters. Digital images were taken using a Hamamatsu ORCA II camera (Hamamatsu, Japan) operated via the SmartCapture Software (Digital Scientific, UK).

Human and mouse BAC arrays

As control experiments both human and mouse BAC arrays have been used. BACs for the 5 K human array were collected and amplified as described (6) and spotted in triplicate using an OmniGrid® 100 microarrayer (Genomic Solutions, Ann Arbor, MI, USA) equipped with SMP3 pins (TeleChem International, Sunnyvale, CA, USA). After printing slides were processed and blocked prior to use according to the manufacturer's protocol, hybridized as described (6) and scanned using a Microarray Scanner G2505B (Agilent Technologies). Results were analysed as described earlier for oligo arrays with the addition that triplicates from the human BAC arrays were fused by BlueFuse in a weighted fashion. The human BAC clones were mapped to the human genome build NCB134. Production, DNA isolation, DNA labelling, hybridization, scanning and analysis of 1 Mb mouse BAC arrays were done completely according to Chung *et al.* (15). The lambda parameter setting for BAC in ArrayCGHsmooth (14) was set to 1.5.

RESULTS

Performance, sensitivity and reproducibility

We have optimized the quality of our in-house oligo array CGH platform (6) using enhanced, automated hybridization conditions and a newly designed extended oligonucleotide library. In addition, we have used BlueFuse feature extraction software enabling fully automated spot analysis. Using these new measures two major goals were achieved owing to the optimized protocol. First, a significant reduction in the amount of background signal was obtained, resulting in a substantial decrease in the number of flagged spots. Second, the variation as defined by the SD in areas without detected copy number changes was considerably reduced. Simultaneously, we have cut the total hybridization costs by removing the prehybridization and reducing the amount of Cot-1 DNA by 90% without compromising quality.

The capability of the current oligo platform to detect and map chromosomal aberrations is demonstrated by a hybridization of DNA from the breast tumour cell line BT474 with normal male human reference DNA (Figure 1A). In another experiment we hybridized DNA from an individual normal female sample with DNA from a pool of normal male samples (Figure 1B). The genome-wide profiles from these experiments are displayed with a weighted moving average of 250 kb to facilitate visual interpretation. The BT474 profile shows many well-defined gains, losses and amplifications as calculated by the smoothing algorithm (14) and the female–male profile shows a straight line with only the X-chromosome gain and Y-chromosome loss, confirming the sex mismatch. With 25 388 (BT474) or 25 030 (female–male) uniquely mapped oligonucleotides that passed quality control, from a total of 26 845 oligonucleotides, >93% of spots were used for calculations resulting in a resolution of ∼100 kb on average. Reproducibility of these results is demonstrated by three additional technical repeats of the BT474 and seven additional biological repeats of female–male hybridizations. Statistics for these and all other array experiments described are provided in Supplementary Table 1.

Variation between consecutive oligonucleotides in an area without chromosomal aberrations is an accumulative consequence of technical noise and biological variation, which is the copy number variations across individuals (16), in this case BT474 and the pooled reference. Snijders *et al*. (3) make the same distinction for BAC arrays. Technical noise here is defined as the SD of a particular oligonucleotide across a series of experiments, further referred to as intra-oligonucleotide variation. Individual copy number variations cannot be measured independently, but are part of the inter-oligonucleotide noise, defined as the SD within one array and chromosomal area without copy number changes. To estimate the inter-oligonucleotide variation in the BT474 profile we measured the SD of the log2ratio on chromosome 2 where no chromosomal copy number changes are observed. This SD is 0.25 when calculated on the raw data and 0.13 when a weighted moving average with a window of 250 kb is applied. By comparison, the SD on the log2ratios in chromosome 2 in the BT474 BAC profile from Snijders *et al*. (3) (hybridization 4001024) is 0.17. In order to visualize the actual variation and resolution detailed views of chromosome 17 from both experiments without moving average are shown in Figure 1C (BT474) and D (female–male). The amplifications are well-defined and the results suggests a fifth amplification at 17q24.1 (Figure 1C, arrow) that has not been observed in a 1 Mb BAC array (3).

The performance of the current oligo array CGH platform as defined by the capability to detect single copy number differences, was assessed by a hybridization of DNA from the male cell line GM01750 with two validated single copy gains on

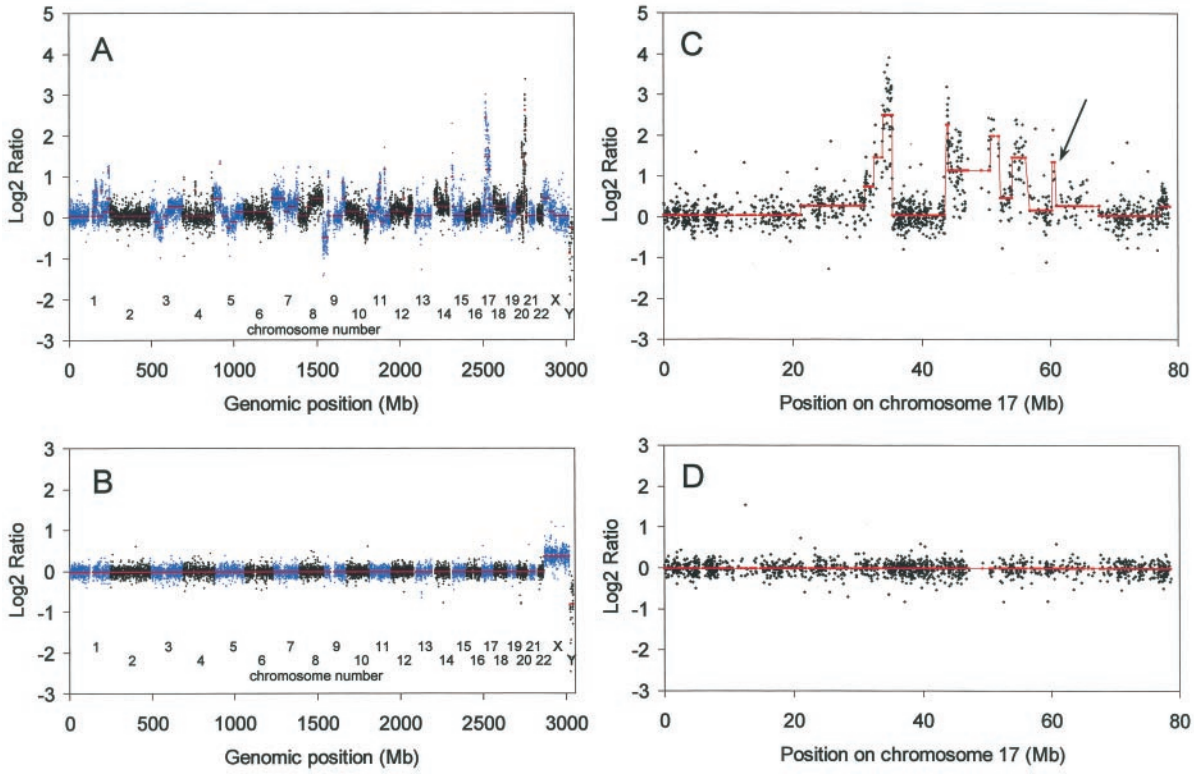

Figure 1. Capability of the oligo array CGH platform to detect and map chromosomal aberrations. Genome-wide profiles are shown that were obtained from hybridization of BT474 DNA (**A**) with human male reference DNA or (**B**) normal male with normal female DNA on a 29 K human oligonucleotide array. Log2ratios were calculated with a weighted moving average as described (11) using a window of 250 kb and are displayed as a function of their position in the genome. Log2ratios of the odd and even chromosomes are shown in aqua blue and black, respectively. Chromosome numbers are indicated. Smoothed values of the log2ratios were calculated using a dedicated smoothing algorithm (14) (red). Note the many breakpoints, gains, losses and amplifications in the BT474 profile and the lack of those in the male–female profile. Detailed profiles of chromosome 17 for the BT474 (**C**) and male–female (**D**) hybridizations. Log2ratios were calculated without moving average and are displayed in black as a function of their position on chromosome 17. Smoothed values of the log2ratios (red). The arrow in C indicates a fifth amplification in the BT474 profile on chromosome 17 that was not observed in a 1 Mb BAC array (3).

chromosomes 9 and 14, respectively (3), and a pool of normal female reference DNA. Median and SD of the log2ratio from the chromosome 9 single copy gain (ratio = 3/2) and double copy X-chromosome variations (ratio = 1/2) are displayed against the theoretical ratios (Figure 2A). The correlation coefficient is 0.98 and the slope is 0.38. Ideally, measured ratios would be the same as theoretical ratios (slope is 1) (1), but alike results with BAC arrays (1) this is not the case. The overlap between the SDs (Figure 2A, blue error bars) indicate that the result of a single oligonucleotide is statistically non-informative to call a single copy number change. If a moving average of 3 is applied to the data (Figure 2A, red error bars), SDs no longer overlap. Thus, three consecutive oligonucleotides are sufficient to call a single copy number change. The single copy gain on chromosome 9 shown in Figure 2B indeed demonstrates that the oligo array CGH platform is powerful for effective breakpoint detection. Similar results were obtained with two more validated cell lines, GM13031 and GM07408 (Supplementary Table 1).

Variation due to intra-oligonucleotide noise was evaluated by analysing four hybridizations using one batch of BT474 DNA with normal male human reference DNA. Hybridizations were done independently on four different days and on three different print batches of the 29 K human oligo arrays. A detailed view of part of chromosome 2 from all four hybridizations shows the resemblance between the profiles from each hybridization (Figure 3). The average of the SDs calculated per oligo (technical noise) from the four experiments is 0.14, whereas for BAC arrays a value of 0.08 was obtained (3). Thus, this experiment shows that variation between the log2ratios obtained from different oligonucleotides within one experiment is reproducible, as shown by the low technical noise value (intra-oligonucleotide noise).

Detection of deletions

The most difficult events to show by array CGH are chromosomal deletions (9). The signal in the test channel is very low, thus can only be accurately measured if background is minimal. Here we show that our current oligo array CGH platform is able to detect both heterozygous and homozygous deletions with high resolution. A potential homozygous deletion was revealed on chromosome 13 in the breast cancer cell line MDA-MB-468 (Figure 4A). The deletion spanned ~100 kb from position 47 883 to 47 977 kb which falls within 13q14.2 encompassing retinoblastoma 1 (RB1) and was not detected in another cell line, illustrated by SUM159 (Figure 4C). To verify the deletion, we applied FISH using a probe for RB1 combined with chromosome 13 paint. We

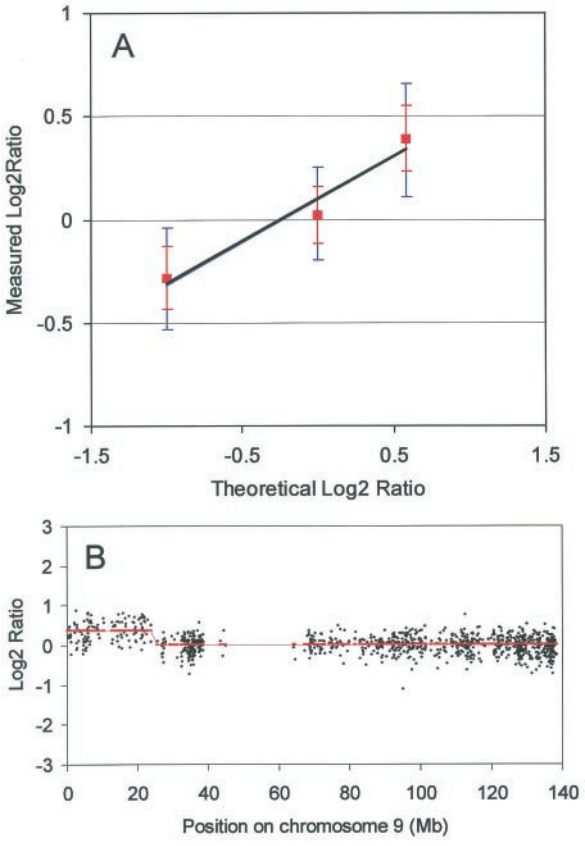

Figure 2. Performance of the oligo array CGH platform. (**A**) DNA from the cell line GM01750 was hybridized against normal female reference DNA. Copy number changes were detected using the smoothing algorithm (14). Blue: median values and SDs (error bars) were calculated for the areas with different copy numbers and are displayed as a function of the theoretical log2ratio. The different areas were the X-chromosome (theoretical ratio = 1/2, log2ratio = −1), chromosomes 1–8, 10–13 and 15–22 (theoretical ratio = 2/2, log2ratio = 0) and the gain in chromosome 9 (theoretical ratio = 3/2, log2ratio = 0.58). The correlation coefficient is 0.98 and the slope is 0.38. Red: same values calculated after applying a moving average of 3 to the data. Note that the red error bars do not overlap. (**B**) Detailed profile of chromosome 9 of the GM01750 hybridization. Log2ratios were calculated without moving average and are displayed in black as a function of their position on chromosome 9. Smoothed values of the log2ratios (red).

found two copies of RB1 (green signals) in metaphases from SUM159 (Figure 4D), but none in MDA-MB-468 (Figure 4B). In addition, chromosome painting showed that there was one normal chromosome 13 and two marker chromosomes with chromosome 13 material. This confirms the homozygous deletion detected by oligo array CGH.

Another potential deletion was detected on chromosome 12 in the hypo diploid breast cancer cell line SKBR7 (Figure 5A). The oligo ratios indicated a deleted region spanning ~2 Mb. To verify the deletion we applied FISH using 3 BACs that were encompassed within the deleted region: RP11-340F14, RP11-44F24 and RP11-7M8 (Figure 5B). We found an interstitial deletion involving all three BACs on one of the two copies of chromosome 12 in this cell line, confirming the presence of a heterozygous deletion. Thus, both homozygous and heterozygous (micro-) deletions can be detected.

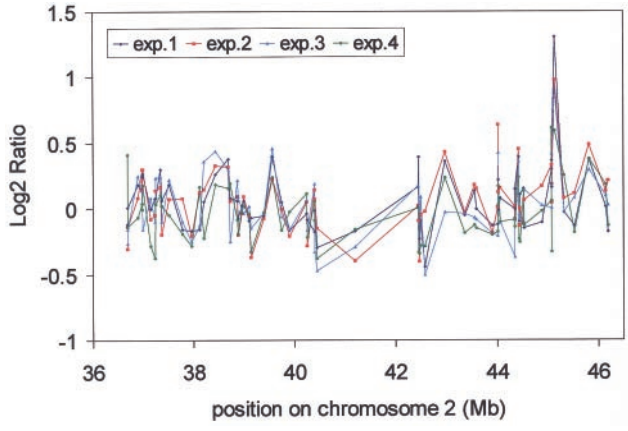

Figure 3. Reproducibility of the oligo array CGH platform. BT474 DNA was hybridized four times against normal human male reference DNA on four different days and on three different batches of 29 K human oligo arrays. Log2ratios were calculated without moving average and are displayed in different colours as indicated for each experiment as a function of their position on chromosome 2.

Performance using DNA from archival specimens

Experiments described above have been performed using high quality DNA obtained from either cell lines or frozen tumour tissue. DNA isolated from FFPE tumour samples is slightly degraded and gives sub-optimal results on BAC array CGH platforms (17). Since the use of FFPE material is invaluable for cancer research, we tested if DNA obtained from FFPE samples can also be used with oligo arrays, which to our knowledge has not yet been reported. DNA obtained from an FFPE gastric tumour was hybridized with normal human reference DNA to a human oligo array and for comparison, with a 1 Mb human BAC array. Figure 6 shows a detailed view of the log2ratios obtained from either the oligo array (Figure 6A) or the BAC array (Figure 6B) for chromosomes 19–21. The genome-wide profile of these as well as technical replicates of these experiments are available via GEO series accession GSE3264 and Supplementary Table 1 provides statistics on the reproducibility. There is a remarkable similarity in the profiles calculated by the smoothing algorithm from the oligo and the BAC array. The amplification in chromosome 19 is clearly detected as well as several levels of deletions in chromosome 21. In addition using the oligo array, the positive log2ratios for amplifications and the negative log2ratios for deletions were higher when compared with the BAC results. Additional FFPE DNA samples as well as lung tumour DNA samples of poor quality due to microdissection were compared using both the oligo and BAC array CGH platforms, yielding matching results (data not shown). Thus, the oligo array CGH platform can perform at least as well as a 1 Mb BAC array platform when it comes to detecting copy number changes in FFPE samples.

Mouse oligo array CGH

The availability of a 1 Mb resolution mouse BAC library (15) provided the opportunity to evaluate the performance of our 21 K mouse oligonucleotide library in array CGH. Genomic DNA obtained from two different mouse tumours was labelled and hybridized to mouse oligo arrays (single printed) and in

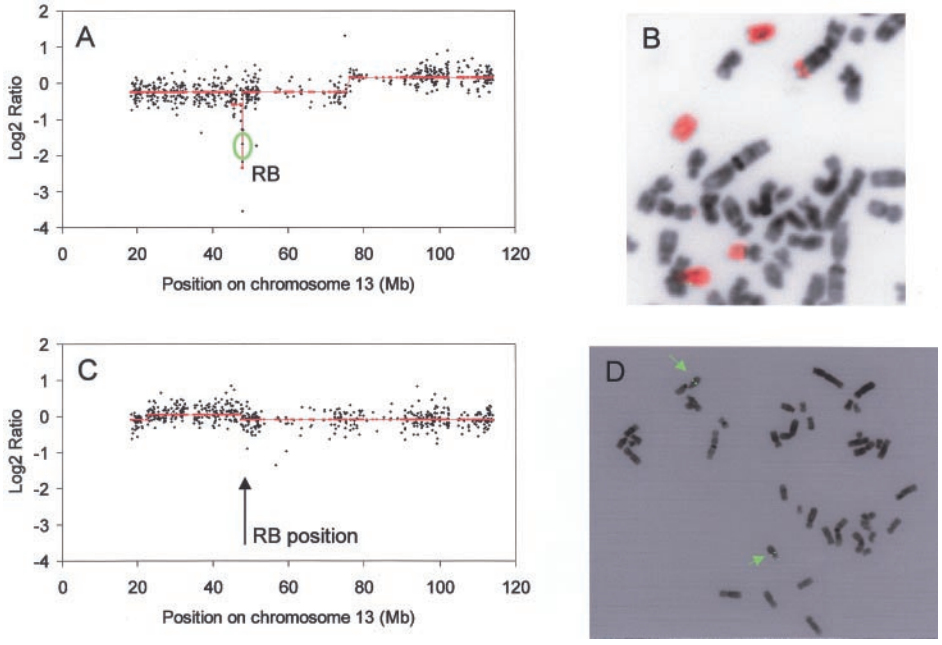

Figure 4. Detection of a homozygous deletion by the oligo array CGH platform. DNA from the cell lines MDA-MB-468 (**A**) and SUM159 (**C**) was hybridized with normal male reference DNA on a human oligo array. Log2ratios were calculated without moving average and are displayed in black as a function of their position on chromosome 13. The smoothed values of the log2ratios are displayed in red and the position of the *RB1* oligo is indicated by the green circle in A and the arrow in C. Note the lack of the deletion in SUM159. Validation of the HD by FISH analysis in cell lines: MDA-MB-468 (**B**) and SUM159 (**D**). The green signal from the *RB1* probe clearly shows the presence of RB1 in SUM159 (D), but is absent from MDA-MB-468 (B). Chromosome 13 paint (red) shows the presence of three normal copies of chromosome 13 and two marker chromosomes with chromosome 13 material.

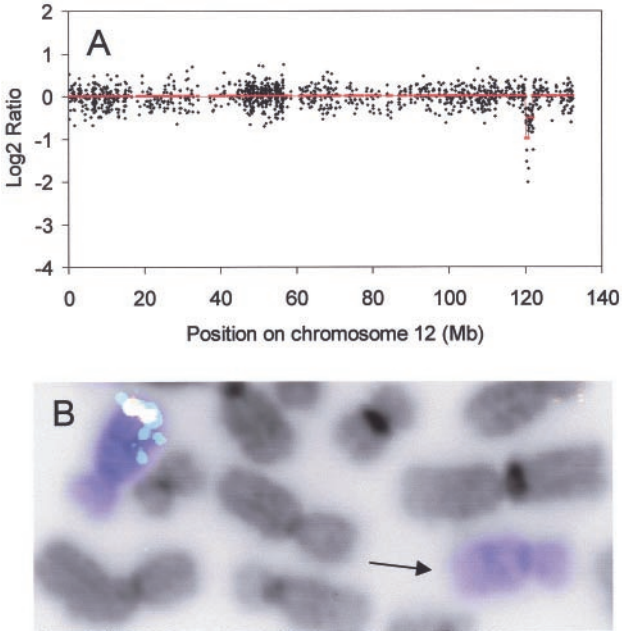

Figure 5. Detection of a heterozygous deletion by the oligo array CGH platform. (**A**) DNA from the cell line SKBR7 was hybridized with normal male reference DNA on a human oligo array. Log2ratios were calculated without moving average and are displayed in black as a function of their position on chromosome 12. The smoothed values of the log2ratios are displayed in red. (**B**) Validation of the heterozygous deletion on 12q24 by FISH analysis of cell line SKBR7. Chromosome 12 paint (dark blue) shows two copies of seemingly normal chromosome 12. FISH using 3 BACs, RP11-340F14 (red), RP11-44F24 (green) and RP11-7M8 (aqua blue) confirmed the interstitial deletion on one copy of chromosome 12 (arrow). Overlapping FISH signals from different BACs show up in white.

fluor-reversed experiments to the mouse BAC arrays (printed in duplicate) (GEO accession GSE3264). Genomic profiles obtained from one of these tumours are shown (Figure 7). All profiles show a striking degree of similarity, despite differences in noise and number of probes. For instance, in Figure 7 chromosome 1 shows two distinct deletions in both profiles and the amplifications as well as the deletion on chromosome 10 completely match. In some cases the smoothed pattern is not exactly the same in both profiles (i.e. chromosome 7), but usually is more refined in the oligo profile, most likely due to the larger number of data points. As observed for the human oligo and BAC arrays in Figure 6 also the positive log2ratios for amplifications and the negative log2ratios for deletions were higher for the mouse oligo arrays compared with the mouse BAC array results. These data show that the oligo array CGH protocol can successfully be applied to mouse oligo libraries in order to assess copy number changes in mouse samples.

DISCUSSION

In this paper we show that the current, optimized protocol for oligo array CGH results in high resolution, reproducible data providing a high quality, cost-effective platform for the detection of chromosomal aberrations in samples from various sources. As our results show, the oligo array CGH platform can detect amplifications (Figure 1), deletions as small as 100 kb (Figure 4), as well as single copy gains (Figure 2) and losses (Figure 5) in single hybridization experiments. When compared with results obtained with a 1 Mb BAC array, the resolution is at least as good and at times

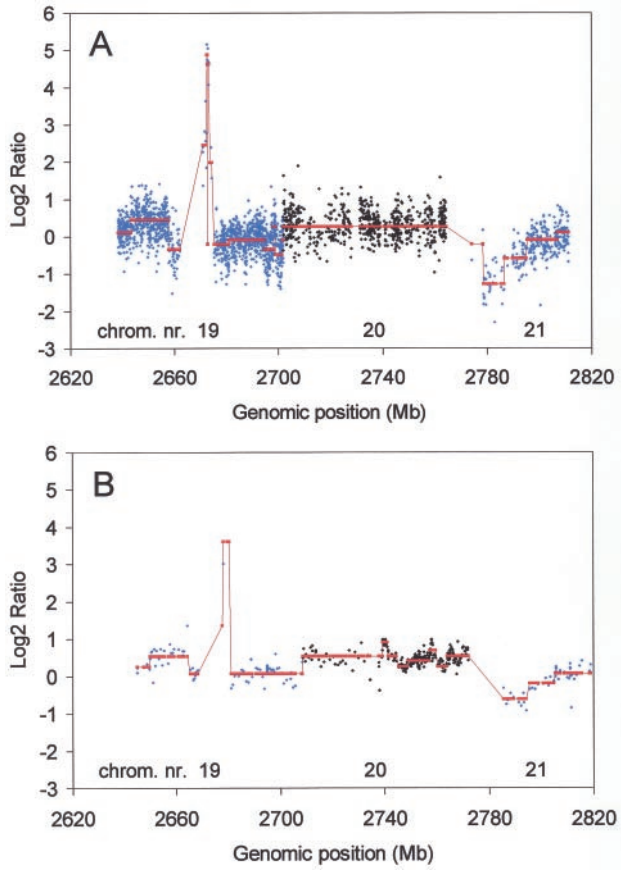

Figure 6. Validation of the oligo array CGH platform with DNA obtained from FFPE tissue. DNA from an FFPE gastric tumour was hybridized with normal human reference DNA on a human oligo array (**A**) and a 1 Mb human BAC array (**B**). Log2ratios were calculated without moving average and are displayed for chromosomes 19–21 as a function of their position on the genome. Log2ratios of the odd and even chromosomes are shown in aqua blue and black, respectively. Chromosome numbers are indicated. Smoothed values of the log2ratios (red).

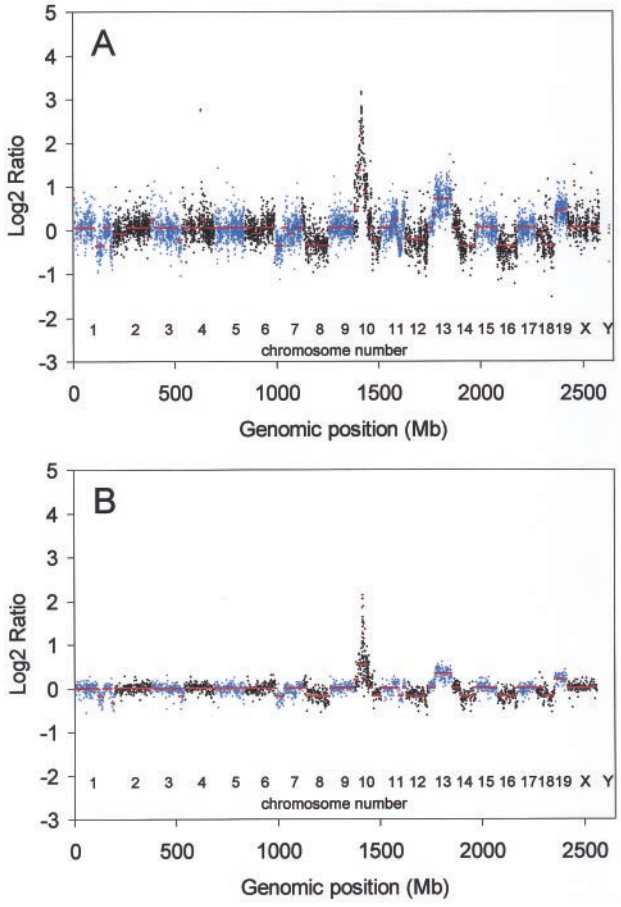

Figure 7. Validation of the mouse oligo array CGH platform. DNA from a mouse tumour was hybridized with normal mouse reference DNA on a 21 K mouse oligo array (**A**). Log2ratios were calculated with a weighted moving average as described (11) using a window of 250 kb and are displayed as a function of their position in the genome. The same mouse tumour and reference DNA was hybridized in a paired fluor-reversed experiment (dye-swap) to a 1 Mb mouse BAC array (**B**). Log2ratios were calculated as described (15) and displayed as a function of their position in the genome. Log2ratios of the odd and even chromosomes are shown in aqua blue and black, respectively. Smoothed values of the log2ratios (red).

considerably better, providing a more refined genomic profile. Furthermore, we demonstrate that good quality results can be obtained with this platform using archival DNA. Finally, labelling DNA from a mouse tumour sample and hybridizing it to the mouse oligo array (Figure 7) showed how easy it was to apply the oligo array CGH platform to a new species.

Optimized hybridization conditions resulted in better signal-to-background ratios. As a result, when using a good quality DNA sample, <5% of spots would not pass quality control, much less than the 65% we previously reported (6). In addition, the inter-oligonucleotide noise as defined by the SD in an area without chromosomal aberrations has been considerably decreased from 0.45 previously (6) to 0.25 with the current protocol (see also Supplementary Figure 1 allowing a direct comparison of the optimized protocol with the previous one). On average from all fresh human DNA samples analysed in this study using the present protocol, the -oligonucleotide noise is even as low as 0.22 (±0.03) with 91% (±5%) of spots passing quality control (data extracted from Supplementary Table 1). Moreover, we show that the technical noise (intra-oligonucleotide variation) is as low as 0.14. Finally, only 300 ng of input DNA is needed for efficient labelling, almost one order of magnitude less than that needed in other protocols (11,18), omitting the need for amplification for many samples.

Breakpoints and copy number changes are easily detected using a dedicated smoothing algorithm (14), although the performance of the oligo array platform, as defined by the relation between the measured and theoretical ratio (11), may not be the most optimal (Figure 2). The platform is also sensitive enough to detect single copy gains (Figure 2) as well as heterozygous deletions (Figure 5). It is likely that homozygous and heterozygous deletions can be distinguished from each other if such difference appears within one experiment. Our results do show evidence that the performance of oligo arrays is comparable with those of BAC arrays (Figures 6 and 7). The suppressed performance of measured versus theoretical ratios is also observed using BAC arrays and is thought to result from incomplete suppression of repetitive sequences (3). This explanation would not be applicable to oligonucleotides, because they are designed within unique genomic sequences. We have successfully tried to take advantage of the lack of repetitive sequences in the oligonucleotide library by

experimenting with lower Cot-1 DNA concentrations during hybridization. Removing 90% of the amount of Cot-1 DNA used for hybridization of BAC arrays did not have any effect on the results (Supplementary Figure 2). Removing all Cot-1 DNA is possible as most aberrations were detected, but showed a suppressing effect on the ratios, especially those in deletions (Supplementary Figure 2). Although theoretically there is no reason for the use of Cot-1 DNA, apparently the presence of small amounts of Cot-1 DNA is yet advantageous in our hands. Interestingly, the use of Human Cot-1 DNA for the mouse oligo array hybridization produced good results.

When evaluating the reproducibility of the oligo array CGH platform, we noticed that the technical noise of an individual oligonucleotide in independent repeats (intra-oligonucleotide variation), is considerably lower than the variation between consecutive oligonucleotides within one experiment (inter-oligonucleotide variation). A considerable difference between technical noise and variation amongst consecutive clones has also been demonstrated for BAC arrays (3), which is a consequence of small chromosomal copy number variations between individuals (16,19). Despite the selection of unique oligonucleotides some cross-hybridization may still occur. However, while array CGH-specific design of oligonucleotides, aimed to minimize cross-hybridization, may improve the performance of an oligo array CGH platform, it does not decrease the variation when compared with oligonucleotides designed for expression arrays (11).

When compared with BAC arrays, signal intensities obtained in oligo arrays are relatively low, explaining part of the noise (SD) observed in profiles from oligo arrays. This is however compensated for by the high number of oligonucleotides when compared with the number of BACs in a 1 Mb BAC array. Moreover, in both human and mouse array experiments ratios obtained by the BAC arrays were slightly compressed compared with the corresponding oligo arrays (Figures 6 and 7). The profiles calculated by the smoothing algorithm (14) are usually very similar between oligo and BAC array experiments from the same sample. In areas where the difference in probe density is high and many aberrations occur the smoothing algorithm may result in different profiles. The high number of oligonucleotides then tends to indicate a more refined pattern of aberrations. A moving average is used to illustrate the sensitivity of an individual element on the oligonucleotide CGH arrays. This shows that three oligonucleotides are sufficient to call for a single copy change. In reality the resolution for downstream analysis of the arrays is higher due to the implementation of sophisticated segmentation algorithms such as ArrayCGHsmooth used here (14,20,21). The average spacing between the oligonucleotides in the human library is ~100 kb, but one should realize that the actual spacing can be very much different. Because most oligo libraries have so far been designed for expression arrays, oligo array CGH will mainly focus on gene-rich areas. Also gene-rich chromosomes, such as chromosome 19, will have a lower average spacing (35 kb) than gene-poor chromosomes, such as chromosome 4 or the X-chromosome (170 kb). Although oligonucleotides are not evenly spread along the genome, in 75% of all oligonucleotides in the human library the distance to the next oligo is <100 kb. Thus, theoretically, the resolution is almost one order of magnitude higher than that of a 1 Mb BAC array for most of the genome.

We have done several hundreds of oligo array CGH experiments in our lab and usually obtain clear profiles. Over 150 of these were breast tumour samples and have confirmed results obtained using Vysis BAC arrays (data not shown). We thus strongly believe that this oligo array CGH protocol provides a robust and reproducible platform. We further evaluated whether our optimized protocol could be replicated in independent laboratories. DNA from the cell line MCF7 was labelled and hybridized to a human oligo array both in our laboratory and by an experienced array CGH group (O. Monni, personal communication) according to the current protocol and very similar results were obtained (not shown). In another experiment DNA from a lung tumour sample was hybridized to the 29 K human oligo array that was printed both on a CodeLink as well as a home-made poly-L-lysine coated slide by a group experienced in oligo arrays (L. Miller, personal communication) resulting in high quality data (Supplementary Figure 3) for both slide types. Although the ratios obtained from the poly-L-lysine slide were compressed when compared with those from the CodeLink slide and the noise was lower on the CodeLink slide, indicating that slide chemistry may be an important factor that contributes to the quality of the data, essentially the same aberrations were detected on both slide types.

Although tiling resolution BAC arrays have been developed (5) and now several commercial oligo array CGH platforms are available (Agilent, Nimblegen), we have shown here that in-house spotting of commercially available oligonucleotide libraries can be a cost-effective way to produce high quality, high resolution results. There is no need for expensive and laborious probe amplifications and as oligo libraries for new organisms become available, these can now easily be implemented for array CGH as shown by us for the mouse oligo array CGH platform, which has already proven its value (22). In contrast, to our knowledge only several groups have invested in producing a genome-wide mouse BAC array (15,23). Moreover, when new libraries are designed, it is worth taking into account that they can also be used for array CGH as is already being done by Illumina (www.illumina.com/products/dna/genomesets/meebo_mouse.ilmn). Special care should be taken to design probes that are unique and show minimal cross-hybridization. Finally, because oligonucleotide probes are exactly defined, in contrast to PCR amplified BAC probes, they are more suitable for use in diagnostic applications.

SUPPLEMENTARY DATA

All microarray data described in this paper are available from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/projects/geo/) through series accession GSE3264. Supplementary Table 1 provides an overview of the GEO accession numbers for the individual array experiments.

ACKNOWLEDGEMENTS

The authors would like to thank Nir Gamliel (Compugen USA, Inc) and Tony Cox, (Sanger Institute, UK) for making the oligo sequences publicly available and mapping them onto the human and mouse genomes in Ensembl (http://www.ensembl.org/). We are also grateful to Prof Outi Monni (University of

Helsinki, Finland) and Prof Lance Miller (Genome Institute of Singapore, Singapore) for collaborating and testing the platform and to Dr Mark van der Wiel for statistical advice. We would like to thank Dr Stephen P Ethier for the SUM 159 cell line and Dr Mike O'Hare for the SKBR7 cell line. S.F.C. and C.C. are funded by the Cancer Research UK. We also acknowledge support by the COST B19 action WG4 for S.F.C. This study was supported by the Centre for Medical Systems Biology (CMSB), a centre of excellence approved by the Netherlands Genomics Initiative/Netherlands Organisation for Scientific Research (NWO) and by the sixth framework programme (EU). Funding to pay the Open Access publication charges for this article was provided by VUmc intramural funding.

Conflict of interest statement. None declared.

REFERENCES

1. Pinkel,D., Segraves,R., Sudar,D., Clark,S., Poole,I., Kowbel,D., Collins,C., Kuo,W.L., Chen,C., Zhai,Y. *et al*. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet*., **20**, 207–211.
2. Oostlander,A.E., Meijer,G.A. and Ylstra,B. (2004) Microarray-based comparative genomic hybridization and its applications in human genetics. *Clin. Genet*., **66**, 488–495.
3. Snijders,A.M., Nowak,N., Segraves,R., Blackwood,S., Brown,N., Conroy,J., Hamilton,G., Hindle,A.K., Huey,B., Kimura,K. *et al*. (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nature Genet*., **29**, 263–264.
4. Kallioniemi,A., Kallioniemi,O.P., Sudar,D., Rutovitz,D., Gray,J.W., Waldman,F. and Pinkel,D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, **258**, 818–821.
5. Ishkanian,A.S., Malloff,C.A., Watson,S.K., DeLeeuw,R.J., Chi,B., Coe,B.P., Snijders,A., Albertson,D.G., Pinkel,D., Marra,M.A. *et al*. (2004) A tiling resolution DNA microarray with complete coverage of the human genome. *Nature Genet*., **36**, 299–303.
6. Carvalho,B., Ouwerkerk,E., Meijer,G.A. and Ylstra,B. (2004) High resolution microarray comparative genomic hybridization analysis using spotted oligonucleotides. *J. Clin. Pathol*., **57**, 644–646.
7. Ren,H., Francis,W., Boys,A., Chueh,A.C., Wong,N., La,P., Wong,L.H., Ryan,J., Slater,H.R. and Choo,K.H. (2005) BAC-based PCR fragment microarray: high-resolution detection of chromosomal deletion and duplication breakpoints. *Hum. Mutat*., **25**, 476–482.
8. Dhami,P., Coffey,A.J., Abbs,S., Vermeesch,J.R., Dumanski,J.P., Woodward,K.J., Andrews,R.M., Langford,C. and Vetrie,D. (2005) Exon Array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am. J. Hum. Genet*., **76**, 750–762.
9. Pinkel,D. and Albertson,D.G. (2005) Array comparative genomic hybridization and its applications in cancer. *Nature Genet*., **37**, S11–S17.
10. Brennan,C., Zhang,Y., Leo,C., Feng,B., Cauwels,C., Aguirre,A.J., Kim,M., Protopopov,A. and Chin,L. (2004) High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res*., **64**, 4744–4748.
11. Barrett,M.T., Scheffer,A., Ben-Dor,A., Sampas,N., Lipson,D., Kincaid,R., Tsang,P., Curry,B., Baird,K., Meltzer,P.S. *et al*. (2004) Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc. Natl Acad. Sci. USA*, **101**, 17765–17770.
12. Winterberg,K.M., Luecke,J., Bruegl,A.S. and Reznikoff,W.S. (2005) Phenotypic screening of *Escherichia coli* K-12 Tn5 insertion libraries, using whole-genome oligonucleotide microarrays. *Appl. Environ. Microbiol*., **71**, 451–459.
13. Weiss,M.M., Hermsen,M.A., Meijer,G.A., van Grieken,N.C., Baak,J.P., Kuipers,E.J. and van Diest,P.J. (1999) Comparative genomic hybridization. *Mol. Pathol*., **52**, 243–251.
14. Jong,K., Marchiori,E., Meijer,G., Vaart,A.V. and Ylstra,B. (2004) Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics*, **20**, 3636–3637.
15. Chung,Y.J., Jonkers,J., Kitson,H., Fiegler,H., Humphray,S., Scott,C., Hunt,S., Yu,Y., Nishijima,I., Velds,A. *et al*. (2004) A whole-genome mouse BAC microarray with 1-Mb resolution for analysis of DNA copy number changes by array comparative genomic hybridization. *Genome Res*., **14**, 188–196.
16. Sebat,J., Lakshmi,B., Troge,J., Alexander,J., Young,J., Lundin,P., Maner,S., Massa,H., Walker,M., Chi,M. *et al*. (2004) Large-scale copy number polymorphism in the human genome. *Science*, **305**, 525–528.
17. Devries,S., Nyante,S., Korkola,J., Segraves,R., Nakao,K., Moore,D., Bae,H., Wilhelm,M., Hwang,S. and Waldman,F. (2005) Array-based comparative genomic hybridization from formalin-fixed, paraffin-embedded breast tumors. *J. Mol. Diagn*., **7**, 65–71.
18. Pollack,J.R., Perou,C.M., Alizadeh,A.A., Eisen,M.B., Pergamenschikov,A., Williams,C.F., Jeffrey,S.S., Botstein,D. and Brown,P.O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet*., **23**, 41–46.
19. van Ommen,G.J. (2005) Frequency of new copy number variation in humans. *Nature Genet*., **37**, 333–334.
20. Willenbrock,H. and Fridlyand,J. (2005) A comparison study: applying segmentation to array CGH data for downstream analyses. *Bioinformatics*, **21**, 4084–4091.
21. Lai,W.R., Johnson,M.D., Kucherlapati,R. and Park,P.J. (2005) Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics*, **21**, 3763–3770.
22. van de Wiel,M.A., Costa,J.L., Smit,K., Oudejans,C.B., Bergman,A.M., Meijer,G.A., Peters,G.J. and Ylstra,B. (2005) Expression microarray analysis and oligo array comparative genomic hybridization of acquired gemcitabine resistance in mouse colon reveals selection for chromosomal aberrations. *Cancer Res*., **65**, 10208–10213.
23. Snijders,A.M., Nowak,N.J., Huey,B., Fridlyand,J., Law,S., Conroy,J., Tokuyasu,T., Demir,K., Chiu,R., Mao,J.H. *et al*. (2005) Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Res*., **15**, 302–311.