

High-resolution array comparative genomic hybridization of single micrometastatic tumor cells

Christine Fuhrmann¹, Oleg Schmidt-Kittler¹, Nikolas H. Stoecklein¹,
Karina Petat-Dutter¹, Christian Vay¹, Kerstin Bockler¹, Richard Reinhardt²,
Thomas Ragg³ and Christoph A. Klein^{1,*}

¹Division of Oncogenomics, Department of Pathology, University of Regensburg, Franz-Josef-Strauß-Allee 11, 93053 Regensburg, ²Max-Planck Institute for Molecular Genetics, 14195 Berlin and ³Quantiom Bioinformatics GmbH & Co. KG, Ringstr. 61, 76356 Weingarten, Germany

Received January 29, 2008; Accepted February 22, 2008

ABSTRACT

Only few selected cancer cells drive tumor progression and are responsible for therapy resistance. Their specific genomic characteristics, however, are largely unknown because high-resolution genome analysis is currently limited to DNA pooled from many cells. Here, we describe a protocol for array comparative genomic hybridization (array CGH), which enables the detection of DNA copy number changes in single cells. Combining a PCR-based whole genome amplification method with arrays of highly purified BAC clones we could accurately determine known chromosomal changes such as trisomy 21 in single leukocytes as well as complex genomic imbalances of single cell line cells. In single T47D cells aberrant regions as small as 1–2 Mb were identified in most cases when compared to non-amplified DNA from 10⁶ cells. Most importantly, in single micrometastatic cancer cells isolated from bone marrow of breast cancer patients, we retrieved and confirmed amplifications as small as 4.4 and 5 Mb. Thus, high-resolution genome analysis of single metastatic precursor cells is now possible and may be used for the identification of novel therapy target genes.

INTRODUCTION

The concept of a hierarchical organization of solid cancers holds that only few selected cancer cells—possibly cancer

stem cells—give rise to local relapses and distant metastases and are responsible for therapy resistance. Interestingly, the presence of single disseminated cancer cells in bone marrow of breast cancer patients without manifest metastasis confers a high risk for later occurring metastatic relapse (1). These extremely rare cells seem to comprise the metastatic precursor cells. In breast cancer patients, genetic analysis using single cell metaphase CGH revealed that they disseminate very early during genomic progression of the disease (2) and differ from their matched primary tumors (3). Therefore the genetic analysis of primary tumors is apparently unsuited to detect changes that are selected during metastatic dissemination. Since identification of such changes may lead to novel therapy target genes, direct high-resolution analysis of the metastatic precursor cells seems to be mandatory.

Recently, resolution for the detection of DNA copy number changes could be increased by array CGH about 100-fold for DNA from pooled cells (4). Reliable measurements of sub-megabase single-copy number imbalances have been reported for amounts of 1–500 ng of high-quality DNA (4,5). However, application of array CGH to single cells was only possible when the data were smoothed over large regions, a procedure that resulted in a resolution equivalent to metaphase CGH (6,7). We had previously developed a protocol suited for metaphase CGH for whole genome amplification of single cell DNA and for analysis of loss of heterozygosity (2,8). We therefore investigated whether a protocol could be established that allows hybridization of amplified single cell DNA to CGH arrays achieving similar genomic resolution as for pooled DNA. Hybridization to existing BAC- and

*To whom correspondence should be addressed. Tel: +49 941 9446720; Fax: +49 941 9446719; Email: christoph.klein@klinik.uni-regensburg.de
Present address:

Oleg Schmidt-Kittler, Howard Hughes Medical Institute, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins School of Medicine, Baltimore, Maryland 21231, USA

Nikolas H. Stoecklein and Christian Vay, Klinik für Allgemein-, Viszeral- und Kinderchirurgie, Universitätsklinikum Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© 2008 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

oligonucleotide-CGH arrays was compared with hybridization to specifically designed BAC-arrays and identified preparation of the microarray and the quality of the single cell sample as critically important.

MATERIALS AND METHODS

Construction of 3K BAC array

BAC isolation. The FISH-mapped BAC clone library (FISH Mapped Clones V1.3) was purchased from BACPAC Resources Center (<http://bacpac.chori.org/>). All BAC-clones were subjected to T7-end-sequencing, the end-sequences were BLASTed against the human genome and annotation was corrected if necessary. In total 11.2% of the BAC-clones were not annotated correctly, out of those 3% were localized on chromosomes different from the ones indicated.

Bacteria were grown in LB medium containing 20 µg/ml chloramphenicol. Standard preparation of BAC DNA was performed by using Qiagen midi columns (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. For preparation of BAC DNA by pulsed-field gel electrophoresis, bacterial agarose-inserts were prepared as described with minor modifications (9) and digested with *PI-SceI*. Briefly, the bacteria were pelleted by centrifugation, embedded in 1% low-melt agarose and lysed with lysozyme and RNase (over night incubation at 37°C with 6 mM Tris-HCl pH 7.5, 1 M NaCl, 100 mM EDTA, 0.5% Brij-58, 0.2% deoxycholate, 0.5% lauryl sarcosine, 30 U/ml RNase (DNase free), 50 000 U/ml lysozym) followed by proteinase K digestion of the cellular proteins (overnight incubation at 50°C with 0.5 M EDTA pH 9.0, 1% lauroyl-sarcosine, 1.5 U/ml proteinase K). Fragmented DNA was removed by gel-electrophoresis at 4 V/cm for 2 h in 0.5×TAE buffer. The BAC DNA was digested with *PI-SceI* for 3 h at 37°C after allowing the enzyme to diffuse into the agarose-insert by overnight incubation on ice (1.5 U/100 µl *PI-SceI*; 100 mM KCl; 10 mM Tris-HCl pH 8.6; 10 mM MgCl₂; 1 mM dithiothreitol; 3.5 µg BSA). Subsequently, the proteinase K digest was repeated. After inactivation of proteinase K by auto-hydrolysis, the agarose-inserts were equilibrated in TE, loaded on a 1% low-melt agarose gel, and the DNA was separated by PFGE in 0.5×TAE using settings recommended by the manufacturer (Amersham Bioscience). The BAC DNA was gel-extracted by an overnight agarose digest (1 U agarase per 100 µl gel-volume in 30 mM Bis-Tris pH 4.5; 10 mM EDTA) and precipitated with isopropanol and Na-acetate (3M) at 4°C. After washing with 70% ethanol, the DNA was resuspended in 10 µl TE.

Amplification of isolated DNA. All BAC DNA preparations were amplified by the adapter-linker method (8,10) with minor modifications and by DOP-PCR as described (11). To avoid cross-hybridization with the adaptors from the single cell PCR products, we selected different sequences for the adaptors that were ligated to BAC DNA, i.e. 5'-CTGTGTCTGACACTCAGTCT-3' and 5'-TAA GACTGAGTCdd-3'. Phi29 amplifications were performed with the TempliPhi Kit (Amersham Biosciences) according to manufacturer's recommendations.

Preparation of BAC microarray. BAC probes were ethanol precipitated after PCR or strand displacement amplification. For the test array, comprising 30 BAC clones derived from the FISH-mapped clones of the V1.3 library (for clone names and localization see Supplementary Table 1), DNA concentration was adjusted to 0.4 µg/µl in 3×SSC. Probes were printed in replicas of five on amino-silane coated slides (Corning, GAPS II) using an Omnigridd 100 spotting robot (Genomic Solutions, Ann Arbor, USA). DNA was immobilized by UV cross-linking followed by 30 min baking at 80°C.

For the 3K array DNA concentration was adjusted to 0.8 µg/µl in water and complemented with 2×*sciSPOT* spotting-buffer (Scienion, Berlin) to a final concentration of 0.4 µg/µl. The slides were printed in replicas of two and post-processed by Scienion AG, Berlin, Germany, using the *sciPROCESS* buffer. For this, slides were washed in 55°C *sciPROCESS* buffer for 15 min, then washed once in distilled water for 5 min at 55°C and twice in distilled water for 25 min at 25°C. Then the slides were dried by centrifugation at 900 rpm. Spot quality was assessed after staining with Sybr Green.

Labeling of sample DNA

Adapter-linker amplified test and reference DNA were labeled by PCR in presence of Cy5 or Cy3 conjugated nucleotides, respectively. For all single cell array CGH experiments the same male reference DNA was used, which was generated from 1000 cells after *MseI* digest and amplification. The 100 µl PCR mix comprised 13 mM Tris-HCl pH 9.5, 3.25 mM MgCl₂, 5 µM LIB1 primer (10), 200 µM dATP and dGTP, 180 µM dTTP and dCTP, 20 µM Cy3/Cy5-dUTP (Amersham Biosciences) and Cy3/5-dCTP (Amersham Biosciences), 16 U Thermo Sequenase (Amersham Biosciences), 0.8 µl template DNA. The MJ-research thermocycler was programmed to 9 cycles of 94°C 1 min, 60°C 30 s, ramp to 72°C with 0.2°C/s, 72°C 5 min with 60 s extension/cycle.

Non-amplified genomic DNA from T47D cell line and leukocytes of healthy donors and a patient with trisomy 21 were prepared with Wizard Genomic DNA Purification Kit (Promega) according to the manufacturer's recommendations. Labeling of pooled genomic DNA and human placental reference DNA was performed according to Snijders *et al.* (12) using the BioPrime DNA labeling system (Invitrogen). Unincorporated nucleotides were removed by a Sephadex G50 column. For hybridization of the 19 K array two vials of labeled DNA were pooled.

Hybridization of the 3 K BAC array

Eluted labeled DNA was precipitated together with 80 µg C₀t-1-DNA (Roche) and resolved in 45 µl mastermix (50% formamide, 2×SSC, 4% SDS and 0.08 g/ml dextrane (MW>500 000; Fisher Biotech). The samples were allowed to resolve for 1 h at 42°C, were then denatured at 75°C for 10 min and incubated for 2 h at 42°C to allow blocking of repetitive sequences. Hybridization was performed without cover slip at 42°C in a humidified chamber for two nights on a shaker as described (13) for the test array. The 3K array was hybridized with coverslip for 24 h in

a *SlideBooster* (Implen, Munich) at 42°C and a mixing-pausing ratio of 3:7. Before hybridization of target DNA, the array was pre-hybridized overnight at 42°C with 80 µl mastermix as above, containing 250 µg herring sperm DNA and 100 µg mock control DNA (adapter-linker PCR product, amplified without addition of cellular DNA). Following hybridization, slides were washed 10 min in 1 × PBS/0.05% Tween at room temperature, 30 min in 2 × SSC/50% formamide at 42°C and 10 min in 1 × PBS/0.05% Tween (all pH 7.4) at room temperature and finally dried by centrifugation.

Analysis of array data

Arrays were scanned with an Axon 4000A scanner (Axon Instruments, Union City, USA). GenePix Pro V4.0 software (Axon Instruments, Union City, USA) was used to locate features automatically and to measure fluorescence intensities for Cy3 and Cy5 channels.

Data analysis of the 30 BAC clone test array was performed according to Quackenbush (14). Briefly, the geometric means of the BAC clone replicas were calculated and the fluorescence signal intensities of the Cy3 and Cy5 channel were globally normalized. BAC clones in chromosomal areas affected by copy number losses or gains were excluded when the signal intensity was normalized. For all BAC clones, the ratio of the normalized fluorescence intensity was log₂-transformed and plotted against the chromosomal location. Copy number differences were assumed to be significant, if the signal intensity ratios of the affected BAC clones exceeded the upper or lower thresholds defined by the three-times standard deviation that was calculated from the baseline noise. The 3K array data were analyzed using the CAPweb software (<http://bioinfo-out.curie.fr/CAPweb/>) (15) and visualized using VAMP (16). Only features with reference fluorescence intensity at least two-fold over the background were included in the analysis. For all other parameters, default values of the software were used (except for the settings of MANOR.nk which were adapted to the size of the array). For visualization of the data normalized, log₂-transformed fluorescence ratios are plotted against genomic location. All shown hybridization profiles are displayed in the 'gained/lost color code' modus. The breakpoint detection and status assignment of genomic regions is performed by the GLAD package (<http://bioinfo-out.curie.fr/CAPweb/>), for which default parameters recommended in the manual were used for analysis. Outliers are defined as unreliable spots by the software, for their detection the risk factor α is set to 0.001. Data points whose log₂ ratios are within upper or lower $\alpha/2$ -region of the normal distribution are defined as outliers and are not taken into account for the assignment of normal/gained/loss-status to a region. Fourteen BAC-clones were detected as outliers in 40% of analyzed hybridization profiles and were thus excluded from analysis.

Single-cell PCR, metaphase CGH and sequence-specific PCR

Single cell adapter-linker PCR and metaphase CGH were performed as published (3,8,10).

To assess quality of DNA samples from global genome amplification of single cells, the DNA was subjected to PCR for five different markers. Markers were chosen such that they were localized on large MseI-fragments, ranging from 1034 to 1936 bp. PCR was conducted as previously described by Schardt *et al.* (2), template DNA was diluted 1:50 in nuclease-free water. Primer sequences are available upon request.

Quantitative PCR

Quantitative PCR was performed using the LightCycler System (Roche, Mannheim, Germany) and the LightCycler FastStart DNA Master SYBR Green I kit (Roche) according to the manufacturer's recommendation. The forward and reverse primer sequences for the amplification of *E. coli* TRPE genomic sequence and the pBACe3.6 vector sequence were taken from Foreman *et al.* (17), without the additional probe beacon (primer sequences are provided in Supplementary Table 2). The LightCycler Relative Quantification Software (Roche) was used to evaluate the runs. PCR efficiency normalization and a reference sample were included for every run. The total amount of vector and *E. coli* DNA in the samples was calculated in an exponential equation with the amplification efficacy of the vector or *E. coli* PCR reactions (1.92 and 1.93 respectively) as base to the power of the crossing points for each sample pair. The ratio of vector versus *E. coli* DNA was determined and expressed as percentage of contaminating *E. coli* DNA in the samples.

For the analysis of the gains on chromosomes 12p and 14q in single cells (Figure 5), three primer pairs were selected within the *AKAP3* locus (chromosome 12) and three primer pairs within the *AKAP6* locus (chromosome 14; indicated by asterisks in Figure 5A and B). Three primer pairs within balanced regions (*SMS2* locus on chromosome 4 and *BC* and *IL-6* locus on chromosome 7; indicated by arrows in Figure 5A and B) served as controls. Measurements were performed in duplicates. Primer sequences are provided in Supplementary Table 2. Ten normal single cells were used for statistical comparison. The relative quantification was performed as described in Schardt *et al.* (2). Briefly, to assess the normal relative ratios of the regions we tested all six loci in ten single control cells together with the two tumor cells and an additional micrometastatic cell of the same patient and performed a rank-sum test for statistical differences. To estimate the copy numbers of the *AKAP3* and *AKAP6* genes the normalized ratios of target and reference genes of the tumor cells were divided by the normalized ratios of target and reference genes of the control cells.

Statistical analysis

Correlations between array CGH and metaphase CGH experiments and between array CGH hybridization profiles were calculated by bivariate correlation analysis using a two-tailed Spearman-Rho test. Statistical difference between qPCR values of the analyzed tumor cells and control cells was determined by applying a two-tailed Mann-Whitney non-parametric test.

Significance of differences in GC-content were calculated using a Student's *t*-test.

RESULTS

Contamination with bacterial DNA impedes array CGH of single cells

We first tested arrays that had been successfully used to analyze large amounts of DNA from primary tumors (20) and hybridized DNA amplified from single cells of known karyotypes onto these arrays. Unfortunately, we were unable to correctly determine the chromosome copy numbers of the test samples (data not shown). As underlying reason we found that contaminating bacterial DNA is introduced during single cell sample preparation with enzymes that are needed for the amplification of cellular DNA (proteinase K, MseI, T4 DNA ligase and Taq polymerase). This contamination after PCR amplification depends on the ratio of the starting cellular target DNA to contaminating DNA, i.e. it increases with lower cell numbers (Figure 1A). The bacterial nature of this contamination was verified by sequencing (e.g. *E. coli* sequences L-asparagine transporter, tryptophane synthase subunit B, transcriptional repressor of *hyc* and NAD synthetase were identified). We hypothesized that bacterial or plasmid DNA contaminating our single cell preparations hybridizes to bacterial or plasmid DNA present on the BAC array which then causes large signal fluctuations. To test this hypothesis we hybridized labeled mock control samples that had been prepared without the addition of human DNA. Indeed, no signals were detected on metaphase spreads while the intensely fluorescent hybridization pattern on BAC arrays could hardly be distinguished from samples containing DNA amplified from human cells (data not shown). These pilot experiments revealed that in single cell experiments contaminating bacterial DNA has an even greater influence on hybridization outcome than previously recognized in standard experiments (21).

Purification of BAC DNA, array preparation and hybridization conditions

Consequently, we tested whether BAC array platforms free of contaminating DNA enable array CGH of single cells. Thus, we had to establish BAC preparation essentially free of contaminating non-human DNA, to avoid hybridization of labeled non-human DNA in the sample to homologous sequences on the array. We tested various approaches, including long-range PCR, *in vitro* transcription of the BAC inserts, subtractive enrichment of the BAC DNA, and the purification of the BAC clones by pulsed field gel electrophoresis (PFGE) (9). The latter proved to be superior to all other strategies, because it enabled both highest recovery and best purification of the BAC DNA. The RP11-BAC clones used for this purpose contain a recognition site of the homing endonuclease PI-SceI in the pBACe3.6 backbone sequence, which is neither present in the *E. coli* genome nor in mammalian genomes (22). Indeed, selective linearization of the RP11-BAC clones enabled the physical separation from the

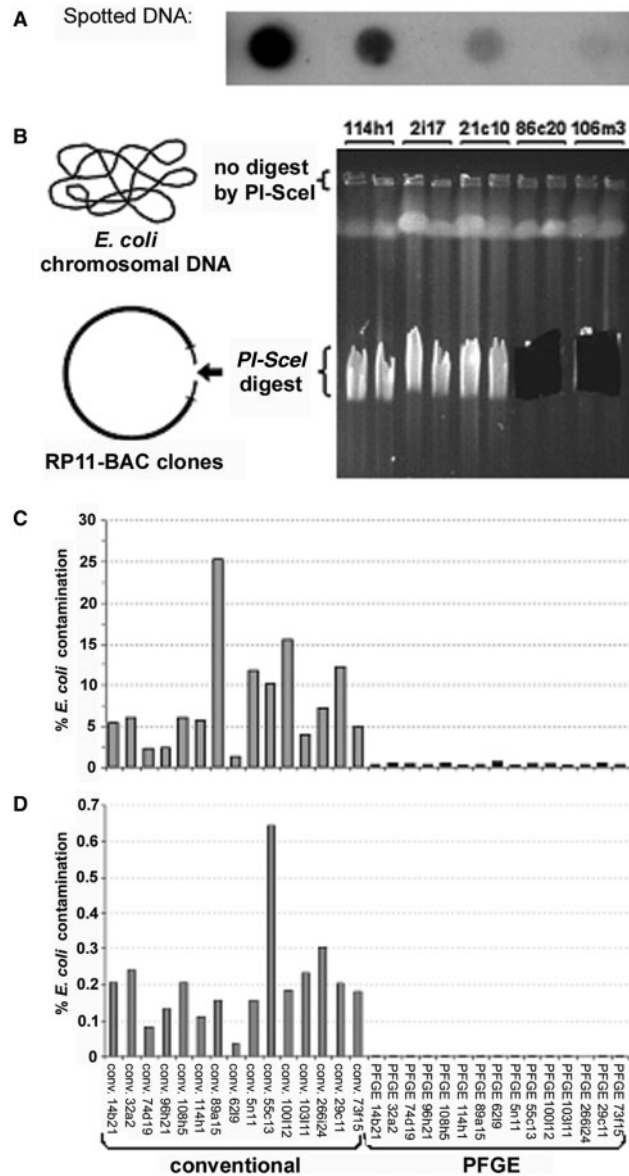


Figure 1. Contamination of amplified DNA samples and PFGE purification of BAC DNA from *E. coli* DNA. (A) Amplified DNA from empty-control for the amplification PCR, 1 cell, 100 cells and 40000 cells (from left to right) were spotted onto a membrane and hybridized with Dig-labeled bacterial contamination. The more human cellular DNA was added, the weaker the hybridization signal of the labeled contamination. The PCR-mock-control contains the highest amount of contamination. (B) Schematic illustration of the pulsed-field-gel-electrophoresis (PFGE) of PI-SceI digested RP11-BAC clones. A PI-SceI site is located in the pBACe3.6 vector sequence, while no restriction site is present in the genomic DNA of *E. coli*. The linearized BAC DNA is recovered after gel-electrophoresis by gel extraction (BAC clones 86c20, 106m3). (C) and (D) *E. coli* contamination of BAC clone preparations determined by qPCR before and after adapter-linker-amplification, respectively. The same 15 BAC clones were prepared by PFGE- and conventional purification. The contamination with *E. coli* DNA is expressed as percentage of *TRPE* sequence in comparison to pBACe3.6 sequence.

circular *E. coli* genomic DNA during PFGE (Figure 1B), whereas circular BACs could not be separated under the applied conditions. The purification of the BAC DNA from *E. coli* DNA was verified by quantitative PCR

(qPCR) using primers for the *E. coli* TRPE gene (17). While we detected only traces of contaminating *E. coli* DNA in PFGE-purified samples ($0.32\% \pm 0.11$ SD), BAC clones conventionally prepared using Qiagen columns were contaminated to a varying degree (1.3–25.2%, average $8.0\% \pm 6.0$ SD) (Figure 1C), the average purity of these column-purified BAC preparations being similar to reported data (6–10%) (17). Also, after amplification of the isolated BAC-DNA, the contamination of column-purified BAC-DNA was substantially higher as compared to PFGE-purified BACs (Figure 1D).

The DNA from 30 RP11 clones, isolated by both PFGE and conventional columns, was used for the generation of a test array. The two BAC DNA preparations were amplified using adapter-linker-PCR, DOP-PCR and Phi29 rolling-circle-amplification, methods which have been previously described for array preparation after conventional BAC preparation (4,12,21,23). The resulting six different DNA samples obtained from each RP11 clone were spotted together onto one glass slide (five replicas each) for a direct comparison of hybridization results.

We first hybridized 500 ng of non-amplified DNA extracted from peripheral blood cells of a patient with trisomy 21 to the test array. All BAC clones from chromosome 21 displayed the expected copy number gain regardless which preparation and amplification method was used (Figure 2A–F), verifying the suitability of both PFGE and conventional BAC DNA isolation methods for this application. However, when DNA from a single leukocyte of the same patient was amplified by adapter-linker PCR (8) and subsequently hybridized onto the differently prepared BAC clone sets (Figure 2G–L; color switch experiments), the trisomy 21 was only detected by PFGE purified clones (Figure 2G–I) but not by column-purified clones (Figure 2J–L). Array CGH with DNA from single leukocytes of healthy donors ($n = 3$) showed normal ratios for all BAC clones (data not shown). Of the three methods for BAC DNA amplification adaptor-linker PCR was superior to DOP-PCR and Phi29-amplification (Figure 2G and M), because it showed the lowest values for false-positive and false-negative signals (Figure 2M). This is possibly due to perfectly matching DNA fragments when probe and target are generated by MseI digest.

Genome-wide array CGH of single cells

With these results in hand we purified 2940 FISH-mapped BAC clones by PFGE and generated an array with an average resolution of 1 Mb. For the validation of this 3 K BAC-array we performed experiments with DNA from samples with known karyotypes to determine hybridization performance. First, \log_2 ratios for balanced BAC clones were found to be -0.01 ± 0.15 for non-amplified DNA ($n = 4$) and -0.02 ± 0.22 for amplified single-cell DNA ($n = 5$). Second, repeated hybridization of identical samples from pooled DNA ($n = 3$) and single-cell samples ($n = 3$) resulted in correlation coefficients of 0.91 ± 0.06 , and 0.84 ± 0.1 , respectively. Suitability of the single cell samples for array CGH analysis was assessed by performing a PCR for 5 markers localized on long

MseI-fragments with sizes of >1000 bp (data not shown). Samples were used for labeling and hybridization when at least 4 of the 5 markers could be detected.

Next, we hybridized single leukocytes from healthy male (data not shown) and female donors (Figure 3A), and from a male patient with trisomy 21 (Figure 3B) against a male reference DNA which was generated from 1000 cells after MseI digest and amplification. The male–male profile displayed balanced clones throughout the genome, the female–male profile showed normal autosomal clones and a relative loss of chromosome Y and gain of chromosome X. In the single cell DNA sample of the patient with trisomy 21, the gain of chromosome 21 was detected readily.

However, chromosome 19 was reproducibly deleted in all analyzed single cell hybridizations. As potential reason we identified the GC-content of BAC clones which was significantly higher in BAC clones of chromosome 19 as compared to chromosome 3 clones (Figure 3C). BAC clones with high GC-content—the critical threshold being about 45% GC-content—were erroneously assigned as deleted. Removing all BAC clones with GC-content above 45% resulted in a correct classification of chromosome 19 (Figure 3D and E). After correction for chromosome 19, the mean false positive rate of all BAC clones was 0.09%. False-negative assignments, i.e. on chromosome 21 for trisomy 21 cells and gonosomes in mismatch experiments, were never observed.

Resolution of 3 K BAC array and comparison with oligonucleotide arrays

We next analyzed complex genomic imbalances in cancer genomes. Since the hallmark of epithelial cancers is chromosomal instability, we searched for a cell line, which displays little inter-laboratory variation of complex genomic rearrangements, and harbors most rearrangements in most of the individual cells. Of the cell lines tested by metaphase CGH (CAMA, CAL51, MCF-7, SKBR3, T47D; data not shown), T47D was the one to meet both criteria best. In addition, T47D had been previously analyzed on a 3 K (12) and on a high-resolution 32 K BAC array (24) and we expected these data to serve as useful independent reference. Therefore, we hybridized non-amplified DNA from pooled T47D cells and two single-cell samples on our 3 K BAC array. In a direct comparison, the identical samples were hybridized on two commercially available oligonucleotide arrays comprising 244 K (Agilent) and 19 K (Operon) oligonucleotides. All samples were identically labeled but were hybridized using optimized conditions for each platform. Samples for the 3 K BAC array and 19 K array were hybridized in our lab, samples for the 244 K array by the manufacturer. No selection of BAC clones or oligonucleotides, e.g. according to GC-content, was performed in any evaluation.

The data obtained from pooled DNA samples proved that all hybridizations worked well, revealing extensive congruency among the different platforms as well as to profiles of the T47D cell line that were previously published. All hybridizations, including the single cell samples were therefore compared with the profile from

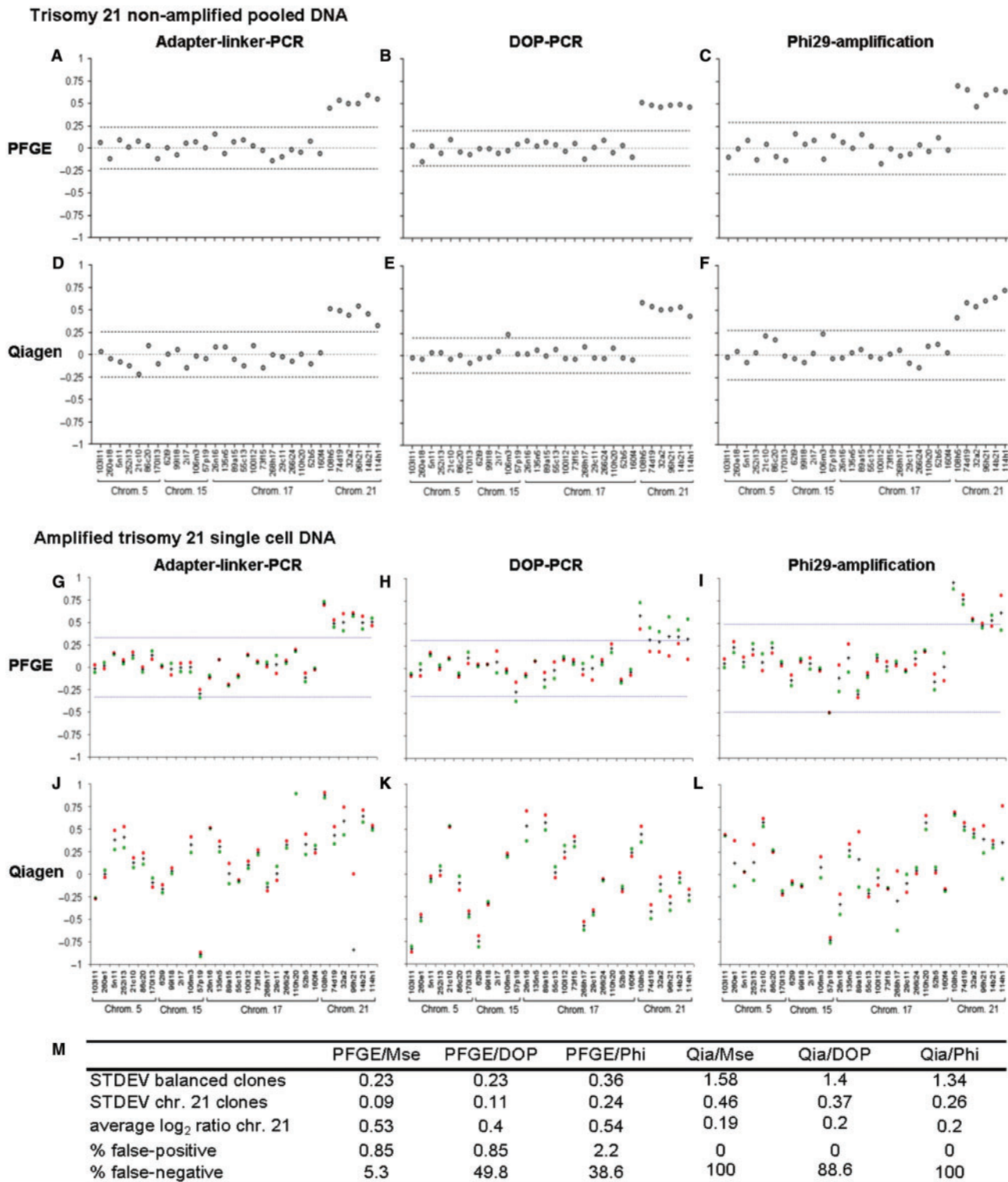


Figure 2. Hybridization of the test-array using DNA from pooled cells and amplified single cell DNA from a patient with trisomy 21. The six different BAC DNA preparations were spotted onto one array and hybridized with Cy3- and Cy5-labeled reference and test DNA probes. The log₂ transformed ratios of test-to-reference DNA fluorescence intensities are plotted against the chromosomal locations of the BAC clones. Dashed gray and black/blue lines represent the base line and the upper/lower significance thresholds defined by the three-times standard deviation, respectively. (A–F) Non-amplified DNA from 10⁶ cells (G–L) Amplified single cell DNA. Note that the trisomy 21 is only detected by the PFGE-amplified BAC-clones (G–I). Red dots represent the experiment in which probe and reference were labeled with Cy5 and Cy3, respectively, the green dots represent the color switch experiment and '+' represents the mean log₂ ratio of both experiments. The 3-fold standard deviation is calculated individually from the balanced clones in each experiment and as threshold the average standard deviation is shown (blue line). (A–C) and (G–I): PFGE purification of BAC clones; (D–F) and (J–L) Qiagen preparation of BAC clones. (A, D, G, J): adapter-linker PCR; (B, E, H, K): DOP-PCR; (C, F, I, L): Phi29 rolling circle amplification. (M) Statistical evaluation of results from different BAC-preparations. Averages from three independent experiments are shown. (Mse: Adaptor-linker-PCR; STDEV: standard deviation).

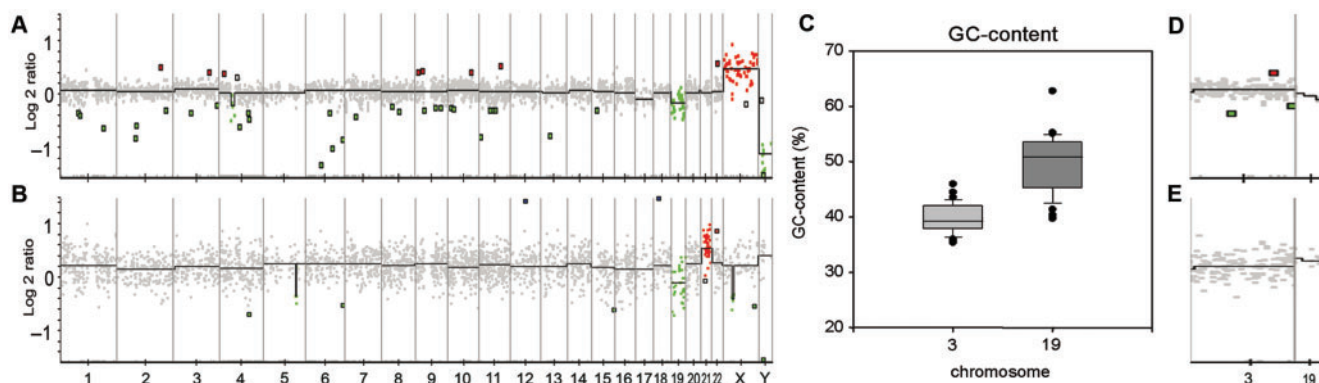


Figure 3. Performance of 3 K BAC array. (A) and (B) Hybridization of amplified single cell-DNA from a female healthy donor and a patient with trisomy 21, respectively, to 3 K BAC array versus male reference DNA. The log₂-transformed fluorescence intensity ratios are plotted against the chromosomal locations of the BAC clones. Gray color of dots indicates normal clones, green indicates loss and red indicates gain of chromosomal material. Blue indicates high-copy amplification. Outliers are marked by a box around the data point; (data normalization and visualization were performed with CAPweb software). (C) GC-content (%) of BAC clones from chromosomes 3 (light gray) and 19 (dark gray) ($n = 30$ clones were analyzed for both chromosomes). (D) and (E) Details of chromosomes 3 and 19 from panels A and B after selection of chromosome 19 BAC clones with GC-content $< 45\%$ ($n = 6$).

pooled DNA on the 244 K array. This allowed comparing directly all platforms by referencing regions of gain and losses to the array with the highest density of features. Thus, we identified normal, gained or deleted regions of different sizes in the reference sample, then determined the BAC clones mapping in these regions, and assessed the percentage of BAC clones that correctly recalled the status of the reference. In all analyses (see Supplementary text and figures), we found that the 3 K BAC array outperformed the oligonucleotide arrays. In particular, detection of small gains and losses (1–20 Mb) was impossible using oligonucleotide arrays (Figure 4). In contrast, on the 3 K BAC array both single cells displayed mostly congruent results with the pooled DNA even for gains and losses as small as 1 Mb (Figure 4 and Supplementary Figure 6). Here, $60.5 \pm 5.2\%$ of gains, losses and interspersed normal regions smaller than 20 Mb were assigned as in the reference DNA. Aberrations and normal regions larger than 20 Mb were concordant for reference and test sample in $77.6 \pm 14.7\%$. Oligonucleotide arrays could rarely retrieve gains and losses and artificially assigned most aberrant regions as being balanced (Figure 4). Identical assignment of small gains, losses and interspersed normal regions to reference was observed for 19 K and 244 K arrays in $40.1 \pm 28.3\%$ and $33.9 \pm 45.7\%$, respectively, and for larger regions (>20 –200 Mb) in $42.8 \pm 35.4\%$ and $36.4 \pm 45.2\%$.

Array CGH of micrometastatic breast cancer cells

Finally, we hybridized individually processed micrometastatic cancer cells that had been isolated from the bone marrow of breast cancer patients (Figure 5 and data not shown). Comparison with metaphase CGH data confirmed the measurements on the array (mean correlation coefficient 0.92 ± 0.02). However, array CGH identified genomic changes that previously escaped detection by metaphase CGH. For example, while on chromosome 11 the loss of the p-arm is readily detected by both methods, the apparent gain of chromosomal material of

chromosome 11q does not become significant in metaphase CGH. As uncovered by array CGH this is caused by two small deleted regions of 7 Mb and 0.5 Mb within an amplified locus—present in both cancer cells—which apparently decreased the averaged signal intensity in metaphase CGH (Figure 5C). Array CGH was able to pinpoint gains on chromosomes 12p and 14q with sizes of 4.4 and 5 Mb, respectively (Figure 5A–C). To validate these amplifications, we applied a previously developed quantitative PCR assay for genomic DNA isolated from single cells (2). Three primer pairs were chosen from each of the amplified regions, comprising the *AKAP3*-locus on 12p and *AKAP6* on 14q (indicated by asterisks in Figure 5A–C). Copy numbers were quantified by comparing qPCR values obtained for primers amplifying the *AKAP3* and *AKAP6* genes to three control loci (*BC* and *IL-6* on chromosome 7 and *SMS2* on chromosome 4; indicated by arrows in Figure 5A and B), which displayed no aberrations. The qPCR assay confirmed the gain of the *AKAP3* and *AKAP6* sequences for all three analyzed tumor cells ($P = 0.001$ for *AKAP3* and $P < 0.001$ for *AKAP6* for each cell) and indicated that the amplicons were present in 4 to 7 copies (Figure 5D).

DISCUSSION

Here we provide a protocol for BAC array CGH of single cells that enabled the detection of copy number changes at high resolution. This was achieved by combining our adaptor-linker PCR amplification protocol for single cells (8) with a novel BAC array using PFGE-purified DNA. This array is essentially free of contaminating bacterial DNA, which was key to success. Experiments using single cells with known karyotypes, such as single normal cells and single cells from patients with trisomy 21, served to test the newly developed 3 K BAC array and yielded the expected results. Complex genotypes of single T47D cells were correctly retrieved using the 3 K BAC array and profiles were concordant with profiles of non-amplified

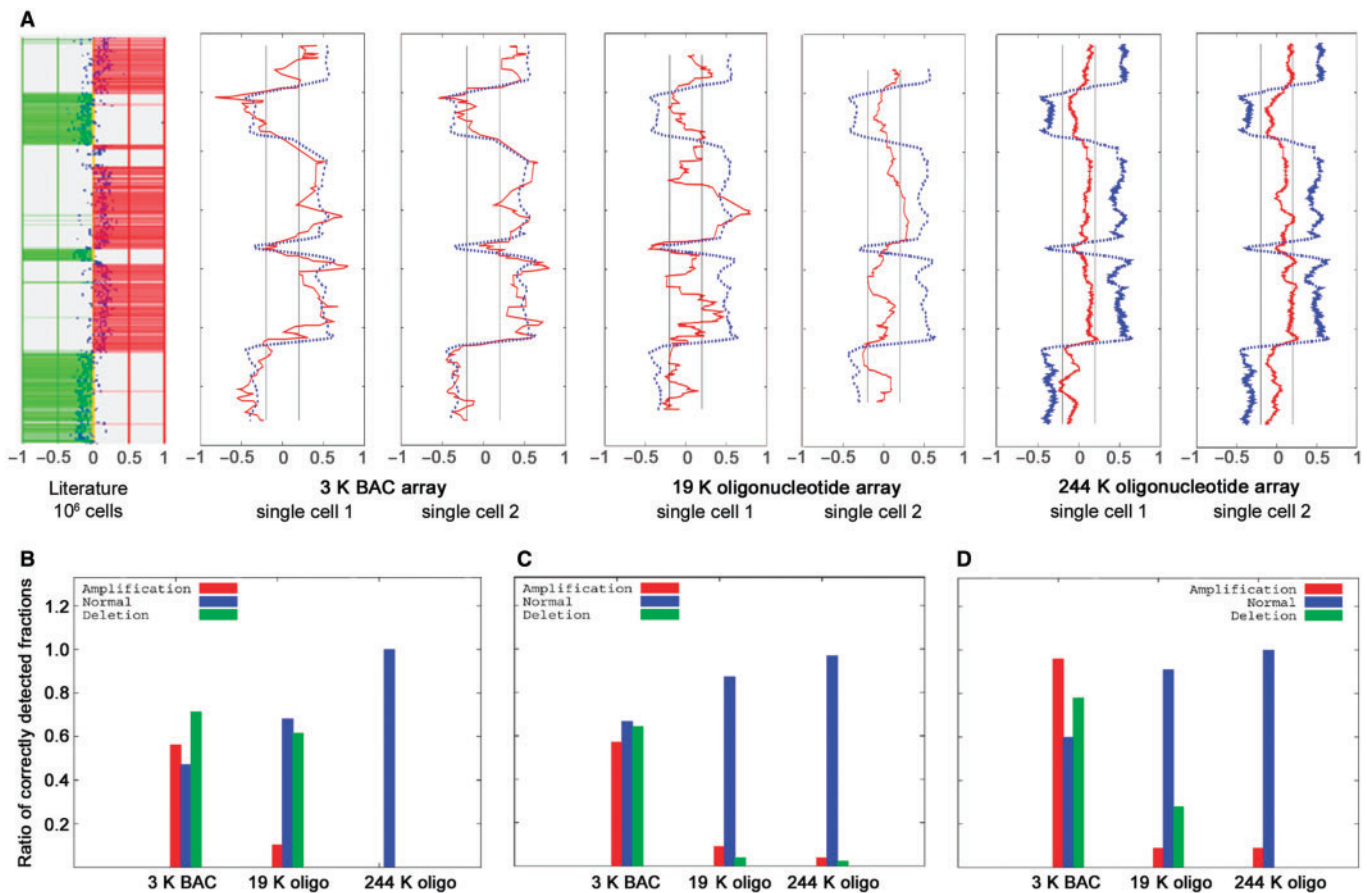


Figure 4. Comparison of the 3 K BAC array and the 19 K and 244 K oligonucleotide arrays. (A) Hybridization profiles of chromosome 10 from hybridization with T47D cell line DNA. From left to right: Hybridization profile of 10⁶ T47D cells to a 32 K BAC array taken from Shadeo *et al.*, hybridization of T47D DNA amplified from two individual single cells on the 3 K BAC array, the 19 K and 244 K oligonucleotide arrays. In all profiles, the blue line represents the hybridization profile of 10⁶ T47D cells on the 244 K oligonucleotide array as a reference. The red line represents the single cell profiles on the various platforms. Log₂ ratios are plotted against chromosomal location of data points. The vertical lines in the profiles indicate the threshold of significant deviations from normal (± 0.2). (B–D) Percentage of correctly detected amplifications, deletions and normal regions in T47D single cell profiles on the three different platforms. Regions between 1 and 20 Mb detected in single cell 1 (B) and single cell 2 (C). Regions between >20 and 200 Mb are provided as an average of single cell 1 and 2 (D).

pooled DNA from T47D cells as well as with published T47D-profiles (24). Comparison of the 3 K BAC array with commercially available 19–244 K oligonucleotide arrays revealed, first, that a data point of the 3 K BAC array is much more reliable than a data point of an oligonucleotide array; second, that a density of 244 000 oligonucleotides is insufficient to perform as good as our 3 K BAC array and third, that single cell analysis on oligonucleotide arrays without knowing the outcome is currently not possible. It remains to be tested in the future whether larger oligonucleotide arrays will enable single cell array CGH analysis.

Although array CGH for single cells has been published previously (6,7,25), the advancement by our approach is underscored by the fact that even a BAC array with 10 times as many BAC clones as ours still did not exceed the resolution of metaphase CGH (6). In contrast, the 3 K BAC array reliably detected gains as small as 4.4 and 5 Mb in single disseminated tumor cells, which were subsequently confirmed by an independent method. This finding is in line with the observation that gains, deletions

and interspersed balanced regions of pooled T47D cells ranging from 1 to 20 Mb were also found in T47D single cells in $60.5 \pm 5.2\%$, including nine regions smaller than 2 Mb. In contrast to the 3 K BAC array, both oligonucleotide arrays were generally unable to correctly assign small and large regions of gains and losses. Since we also observed genomic divergence between pooled T47D cells and the individual cells upon metaphase CGH, some of the gains and losses diverging between reference and single cell samples will be true differences. Therefore, the correct identification of small and large gains or losses by the 3 K BAC array in a single cell is certainly higher than 60% and 77%, respectively.

We obtained first evidence that the performance of our single cell BAC array can be further improved. For example, we already noted that some BAC clones on chromosome 19 led to assignment of incorrect deletions. Adjustment for GC-content of chromosome 19 BAC clones was an adequate remedy. Thus, some BAC clones may not be suited for single cell array CGH and a careful bioinformatic evaluation might uncover rules for

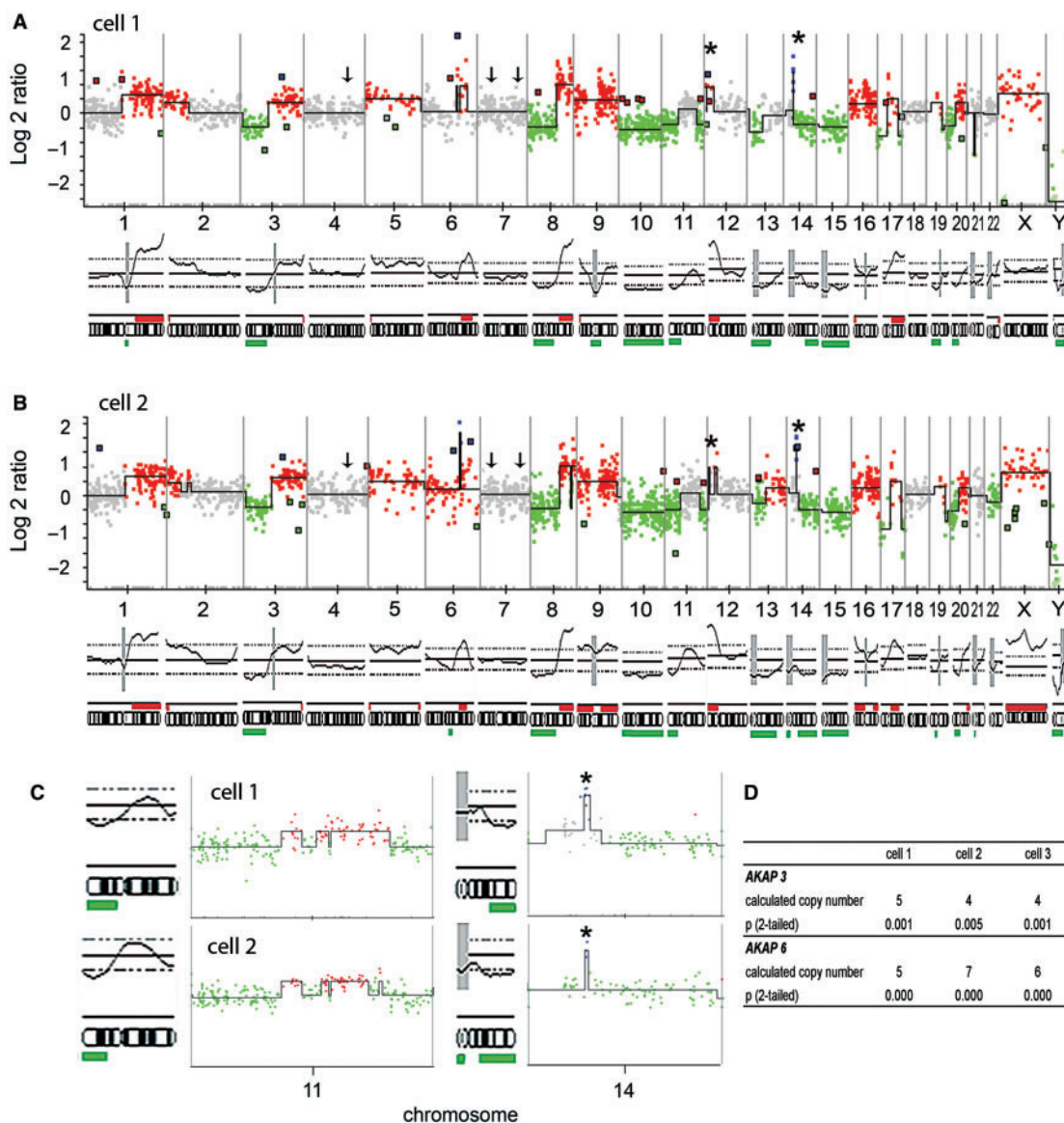


Figure 5. Hybridization of two adapter-linker amplified DNA samples from disseminated tumor cells isolated from bone marrow of a breast cancer patient. The metaphase CGH profile of each sample is plotted beneath the array CGH profile. Chromosome ideograms are placed horizontally to position the corresponding genomic regions of metaphase and array CGH next to each other for direct comparison. (A) Single tumor cell 1, (B) single tumor cell 2, (C) gain of resolution by array CGH. Two small inserted deletions on the q-arm of chromosome 11 (7Mb and 500 kb in size; from centromer to telomer) and the high-copy amplifications on chromosomes 12p (4.4Mb) and 14q (5 Mb) are present in both single cells. Colour code and axes as in Figure 3. Arrows and asterisks point to primer binding sites used as controls or target regions, respectively, for qPCR. (D) Calculated copy numbers of *AKAP3* and *6* in single disseminated tumor cell 1, 2 and 3 as calculated from qPCR results (assuming a diploid genome).

BAC selection. Despite current limitations the first version of our BAC array already enabled the first high-resolution analysis of single micrometastatic cells isolated from bone marrow of breast cancer patients. While the overall profiles correlated significantly with their metaphase CGH profiles, additional changes could be retrieved. These had apparently been selected during cancer progression or were at least expanded within the disseminated tumor cell population as they were found in independently isolated cancer cells from one patient. The concordant results underscore the reliability of the method, as does the quantitative PCR assay for the amplified regions.

The novel array CGH protocol may be suited for numerous additional applications. Examples include the study of minute amounts of *ex vivo* DNA from pre-malignant lesions or very small tumors, or from biopsies before and after systemic therapy. Single cell resolution will also allow addressing questions that could not be asked before. In the current discussion about cancer stem cells the degree of genetic instability can now be assessed in stem cells and transiently amplifying cells to uncover the origin of numerical DNA aberrations. Likewise, genomic causes underlying human ageing (26,27) are thought to be stochastic in nature and notoriously escape when pooled DNA is analyzed. Therefore, array CGH of

single cells will enable novel insights into important biological and medical questions.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was supported by the SFB 456 from the Deutsche Forschungsgemeinschaft, a BioFuture Grant 0311884 from the Federal Ministry of Education and Science, the Wilhelm-Sander Stiftung, and a grant from the Bavarian State Ministry of Sciences, Research and the Arts (to CAK) and a DFG grant STO 464/1-1 (to NHS) and a NGFN-grant 01GR0414/PDN-S02T17 (to RR). We thank Tanja Siart and Barbara Lindner for excellent technical assistance and Dirk Schindlhauer, Evelin Schröck and Bernhard Radlwimmer for their help. Funding to pay the Open Access publication charges for this article was provided by the Bavarian State Ministry of Sciences, Research and the Arts.

Conflict of interest statement. None declared.

REFERENCES

- Braun,S., Pantel,K., Muller,P., Janni,W., Hepp,F., Kentenich,C.R., Gastroph,S., Wischnik,A., Dimpfl,T., Kindermann,G. *et al.* (2000) Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer. *N. Engl. J. Med.*, **342**, 525–533.
- Schardt,J.A., Meyer,M., Hartmann,C.H., Schubert,F., Schmidt-Kittler,O., Fuhrmann,C., Polzer,B., Petronio,M., Eils,R. and Klein,C.A. (2005) Genomic analysis of single cytokeratin-positive cells from bone marrow reveals early mutational events in breast cancer. *Cancer Cell*, **8**, 227–239.
- Schmidt-Kittler,O., Ragg,T., Daskalakis,A., Granzow,M., Ahr,A., Blankenstein,T.J., Kaufmann,M., Diebold,J., Arnholdt,H., Muller,P. *et al.* (2003) From latent disseminated cells to overt metastasis: genetic analysis of systemic breast cancer progression. *Proc. Natl. Acad. Sci. USA*, **100**, 7737–7742.
- 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. *Nat. Genet.*, **36**, 299–303.
- Guillaud-Bataille,M., Valent,A., Soularue,P., Perot,C., Inda,M.M., Receveur,A., Smaili,S., Roest Crolius,H., Benard,J., Bernheim,A. *et al.* (2004) Detecting single DNA copy number variations in complex genomes using one nanogram of starting DNA and BAC-array CGH. *Nucleic Acids Res.*, **32**, e112.
- Fiegler,H., Geigl,J.B., Langer,S., Rigler,D., Porter,K., Unger,K., Carter,N.P. and Speicher,M.R. (2007) High resolution array-CGH analysis of single cells. *Nucleic Acids Res.*, **35**, e15.
- Le Caignec,C., Spits,C., Sermon,K., De Rycke,M., Thienpont,B., Debrock,S., Staessen,C., Moreau,Y., Fryns,J.P., Van Steirteghem,A. *et al.* (2006) Single-cell chromosomal imbalances detection by array CGH. *Nucleic Acids Res.*, **34**, e68.
- Klein,C.A., Schmidt-Kittler,O., Schardt,J.A., Pantel,K., Speicher,M.R. and Riethmuller,G. (1999) Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl. Acad. Sci. USA*, **96**, 4494–4499.
- Smith,C.L. and Cantor,C.R. (1987) Purification, specific fragmentation, and separation of large DNA molecules. *Meth. Enzymol.*, **155**, 449–467.
- Klein,C.A., Blankenstein,T.J., Schmidt-Kittler,O., Petronio,M., Polzer,B., Stoecklein,N.H. and Riethmuller,G. (2002) Genetic heterogeneity of single disseminated tumour cells in minimal residual cancer. *Lancet*, **360**, 683–689.
- Telenius,H., Carter,N.P., Bebb,C.E., Nordenskjold,M., Ponder,B.A. and Tunnacliffe,A. (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics*, **13**, 718–725.
- 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. *Nat. Genet.*, **29**, 263–264.
- 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. *Nat. Genet.*, **20**, 207–211.
- Quackenbush,J. (2002) Microarray data normalization and transformation. *Nat. Genet.*, **32 Suppl.**, 496–501.
- Liva,S., Hupe,P., Neuvial,P., Brito,I., Viara,E., La Rosa,P. and Barillot,E. (2006) CAPweb: a bioinformatics CGH array Analysis Platform. *Nucleic Acids Res.*, **34**, W477–481.
- La Rosa,P., Viara,E., Hupe,P., Pierron,G., Liva,S., Neuvial,P., Brito,I., Lair,S., Servant,N., Robine,N. *et al.* (2006) VAMP: visualization and analysis of array-CGH, transcriptome and other molecular profiles. *Bioinformatics*, **22**, 2066–2073.
- Foreman,P.K. and Davis,R.W. (2000) Real-time PCR-based method for assaying the purity of bacterial artificial chromosome preparations. *Biotechniques*, **29**, 410–412.
- Press, W. (2002) Numerical Recipes in C++. *Cambridge University Press*.
- Pepe, M.S. (2003) The statistical evaluation of medical tests for classification and prediction. *Oxford*.
- Fritz,B., Schubert,F., Wrobel,G., Schwaenen,C., Wessendorf,S., Nessling,M., Korz,C., Rieker,R.J., Montgomery,K., Kucherlapati,R. *et al.* (2002) Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res.*, **62**, 2993–2998.
- Fiegler,H., Carr,P., Douglas,E.J., Burford,D.C., Hunt,S., Scott,C.E., Smith,J., Vetrie,D., Gorman,P., Tomlinson,I.P. *et al.* (2003) DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer*, **36**, 361–374.
- Frengen,E., Weichenhan,D., Zhao,B., Osoegawa,K., van Geel,M. and de Jong,P.J. (1999) A modular, positive selection bacterial artificial chromosome vector with multiple cloning sites. *Genomics*, **58**, 250–253.
- Lage,J.M., Leamon,J.H., Pejovic,T., Hamann,S., Lacey,M., Dillon,D., Segraves,R., Vossbrinck,B., Gonzalez,A., Pinkel,D. *et al.* (2003) Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res.*, **13**, 294–307.
- Shadeo,A. and Lam,W.L. (2006) Comprehensive copy number profiles of breast cancer cell model genomes. *Breast Cancer Res.*, **8**, R9.
- Hu,D.G., Webb,G. and Hussey,N. (2004) Aneuploidy detection in single cells using DNA array-based comparative genomic hybridization. *Mol. Hum. Reprod.*, **10**, 283–289.
- Bahar,R., Hartmann,C.H., Rodriguez,K.A., Denny,A.D., Busuttill,R.A., Dolle,M.E., Calder,R.B., Chisholm,G.B., Pollock,B.H., Klein,C.A. *et al.* (2006) Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*, **441**, 1011–1014.
- Lengauer,C., Kinzler,K.W. and Vogelstein,B. (1997) Genetic instability in colorectal cancers. *Nature*, **386**, 623–627.