High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays

Daniel Pinkel[1,2], Richard Segraves[1], Damir Sudar[2], Steven Clark[1], Ian Poole[3], David Kowbel[2], Colin Collins[2], Wen-Lin Kuo[1], Chira Chen[1], Ye Zhai[1], Shanaz H. Dairkee[4], Britt-marie Ljung[5], Joe W. Gray[1,2] & Donna G. Albertson[1,2,6]

Gene dosage variations occur in many diseases. In cancer, deletions and copy number increases contribute to alterations in the expression of tumour-suppressor genes and oncogenes, respectively. Developmental abnormalities, such as Down, Prader Willi, Angelman and Cri du Chat syndromes, result from gain or loss of one copy of a chromosome or chromosomal region. Thus, detection and mapping of copy number abnormalities provide an approach for associating aberrations with disease phenotype and for localizing critical genes. Comparative genomic hybridization[1] (CGH) was developed for genome-wide analysis of DNA sequence copy number in a single experiment. In CGH, differentially labelled total genomic DNA from a 'test' and a 'reference' cell population are cohybridized to normal metaphase chromosomes, using blocking DNA to suppress signals from repetitive sequences. The resulting ratio of the fluorescence intensities at a location on the 'cytogenetic map', provided by the chromosomes, is approximately proportional to the ratio of the copy numbers of the corresponding DNA sequences in the test and reference genomes. CGH has been broadly applied to human and mouse malignancies. The use of metaphase chromosomes, however, limits detection of events involving small regions (of less than 20 Mb) of the genome, resolution of closely spaced aberrations and linking ratio changes to genomic/genetic markers. Therefore, more laborious locus-by-locus techniques have been required for higher resolution studies[2–5]. Hybridization to an array of mapped sequences instead of metaphase chromosomes could overcome the limitations of conventional CGH (ref. 6) if adequate performance could be achieved. Copy number would be related to the test/reference fluorescence ratio on the array targets, and genomic resolution could be determined by the map distance between the targets, or by the length of the cloned DNA segments. We describe here our implementation of array CGH. We demonstrate its ability to measure copy number with high precision in the human genome, and to analyse clinical specimens by obtaining new information on chromosome 20 aberrations in breast cancer.

The performance goals of array CGH are more stringent than those of related array-based methods for measuring gene expression[7,8]. Optimum utility requires data over a wide copy number range, including reliable detection of single copy changes relative to the normal diploid state. This precision must be achieved for hybridizations involving the entire mammalian genome, a nucleic acid pool that has over ten times the complexity of the expressed sequences, and which includes a significant quantity of interspersed repeats. Our procedures, which differ considerably from others[6–8], meet these challenges when using DNA from cosmid, P1, BAC and other large insert clones for array targets.

The sensitivity and quantitative capability of our methodology was evaluated using test and reference genomes made by adding various amounts of λ DNA (50 kb, 1.5×10^{-5} of the human genome) to the normal human genome. This model simulates the behaviour of a cosmid-sized target that does not contain repetitive sequences. The measured fluorescence ratios on the λ targets were quantitatively proportional to the relative amounts of λ DNA in the test and reference genomes from below single-copy equivalent level to at least two orders of magnitude higher (Fig. 1). Our techniques, therefore, have the sensitivity necessary to measure copy number increases or decreases in the human genome. The limited cross hybridization of human sequences to a λ DNA target is shown (Fig. 3*b*).

Measurement precision and accuracy of array CGH was assessed using arrays containing targets made from a set of clones on chromosome 20 and four clones on the X chromosome. The chromosome 20 target clones (Table 1; Fig. 2*a*, inset) were spaced at an average interval of approximately 3 Mb along the entire chromosome, and included clones specifically chosen to detect regions of copy number change that had previously been found

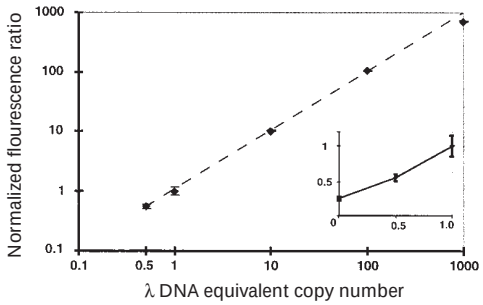

Fig. 1 Comparison of the fluorescence intensity ratio on λ DNA targets to the ratio of λ DNA in model test and reference genomes. The six model test genomes, labelled with fluorescein, contained 400 ng of human genomic DNA and 0, 3, 6, 60, 600 and 6000 pg of λ DNA, respectively. These corresponded to 0, 0.5, 1, 10, 100 and 1000 times the equivalent single-copy level for a 50-kb sequence. The reference genomes for all six hybridizations, labelled with Texas red, contained 400 ng of human genomic DNA and 6 pg of λ DNA. Fluorescence ratios on λ targets were normalized relative to human genomic DNA targets contained in the same arrays. (Cloned targets could not be used for normalization because of cross hybridization between λ DNA and vector sequences.) Ratios for the five test genomes with non-zero λ DNA are shown in the main figure, normalized relative to the 6-pg (single copy equivalent) ratio. Error bars indicate the standard deviations of three independent hybridizations. At some copy number levels the standard deviations are smaller than the symbols used for the data points. The measured fluorescence ratios were quantitatively proportional to copy number over a dynamic range of at least 200, but the ratio was about 30% low at the highest point. The inset shows a linear plot of the data from the 0, 3 and 6 pg hybridizations, 0, 0.5 and 1 times the single copy level, respectively.

[1]*Cancer Genetics Program, UCSF Cancer Center, University of California San Francisco, Box 0808, San Francisco, California 94143-0808, USA.* [2]*Life Sciences Division, E.O. Lawrence Berkeley National laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA.* [3]*Vysis, Inc., Downers Grove, Illinois, USA.* [4]*Geraldine Brush Cancer Research Institute, California Pacific Medical Center, 2330 Clay Street, San Francisco, California 94115, USA.* [5]*Pathology Department, University of California San Francisco, Box 0102, San Francisco, California 94143, USA.* [6]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. Correspondence should be addressed to D.P. (e-mail: pinkel@cc.ucsf.edu).*

Table 1 • Array targets and summary of regional aberrations

Chromosome	Target	Source[a] (library, plate, position)	Locus/gene	Map location[b]	Region[c]	Copy number aberration[d] BT474	S6	S21	S50	S59
Chromosome 20	2005(15)	ref. 5	20p sub-telomere	0.034						
p arm	RMC20P107	DuPont B 114 H10	CDC25B	0.085						
	RMC20P160	DuPont A 967 A8	WI-7829	0.158						
	RMC20P178	DuPont A 899 F11	*D20S186*	0.209						
	RMC20P055	DuPont B 121 B6	*D20S114*	0.272						
	RMC20P099	DuPont B 39 A2	CST3	0.352						
q arm	RMC20P090	DuPont B 81 C11	BCLX	0.526						
	RMC20P117	DuPont B 72 G11	AIB4	0.548	C	●			●	
	RMC20P037	DuPont B 76 F8	*SRC*	0.603	F	○				
	RMC20P154	DuPont A 897 C10	*D20S44*	0.646	F	○				
	RMC20P058	DuPont B 77 D11	*TOP1*	0.675	F	○				
	RMC20P100	DuPont B 53 B2	*SEMG1*	0.694	D	●				
	RMC20P131	DuPont A 213 F1	*D20S178*	0.722	D	●			●	
	RMC20P063	DuPont B 106 D4	*PTPN1/PTP1B*	0.755	B	●			●	●
	RMC20B4135	Research Genetics 22 N5	WI-4444	0.805	A	●	●		●	
	RMC20P4039	DuPont B 34 A6	RMC20C001	0.806	A	●	●		●	●
	RMC20B4097	Research Genetics 133 E8	*D20S211*	0.807	A	●	●	●	●	●
	RMC20B4130	Research Genetics 341 H15	*D20S211*	0.808	A	●	●	●	●	●
	RMC20P070	DuPont B 58 E10	*D20S120*	0.815		●	●			●
	RMC20P071	DuPont B 97 A2	*D20S100*	0.827		●				
	RMC20P073	DuPont B 101 A6	*PCK1*	0.867	E	●	●		●	
	RMC20P179	DuPont A 1173 H1	*CHRNA4*	0.948						

[a]The P1 or BAC clones were selected by PCR from a pooled (DuPont B) or unpooled (DuPont A) human genomic P1 library[29] or human genomic BAC library[30] (Research Genetics). The selection and mapping of many of the P1 clones have been reported[31]. [b]Location of the clones on chromosome 20 are given in fractional length, Flpter previously. Order of the chromosome 20 clones is consistent with genetic maps. [c]Clones that define the regions of consistent copy number aberration are indicated. Those labelled A–E correspond to defined regions[3]; region F is also defined. [d]The filled circles indicate that elevated copy number was measured on that clone, the open circles indicate a loss. All other clones were at average copy number relative to the p arm of chromosome 20.

on chromosome 20q (ref. 3). Four clones were from a contig[9] at 20q13.2 (Table 1, region A). Six independent comparative hybridizations of a normal human genome with itself produced very similar fluorescence ratios on all chromosome 20 targets (Fig. 2*a*), providing a precise baseline from which to measure deviations. The measured ratios, normalized to 1.0 for each hybridization, were 1.0 ± 0.07 (mean ± s.d. for all 6×22 targets equals 132 measurements). The ratios were constant, although the test and reference fluorescence intensities varied several-fold among the targets, approximately proportional to the length of the human DNA segment they contained.

Comparisons of cell populations containing 1–5 copies of the X chromosome with normal female DNA showed that fluorescence ratios were proportional to copy number (Fig. 2*b*). Measurement precision was high enough to resolve each of the four copy number steps. Additional comparisons of normal male and normal female DNA yielded normalized ratios on the X chromosome targets of 0.69 ± 0.05 (mean ± s.d.). Thus, single copy decreases produce ratios that differ from diploid with high statistical significance. Although the measured ratios are proportional to copy number, the slope of the relation, 0.37, differs from the ideal value of 0.5. The reason for this is not known, but reduced slopes are expected if the measured fluorescence intensities have contributions from sources other than unique sequence hybridization[10]. Possible sources include incompletely suppressed repeats (perhaps those not in the Cot-1 fraction), cross-hybridization of closely related sequences and target autofluorescence. Although some repeats may not be completely blocked, suppression of common centromeric sequences is very effective (Fig. 3*b*).

Application of array CGH to breast cancer provided new, high resolution, high dynamic range information on copy number

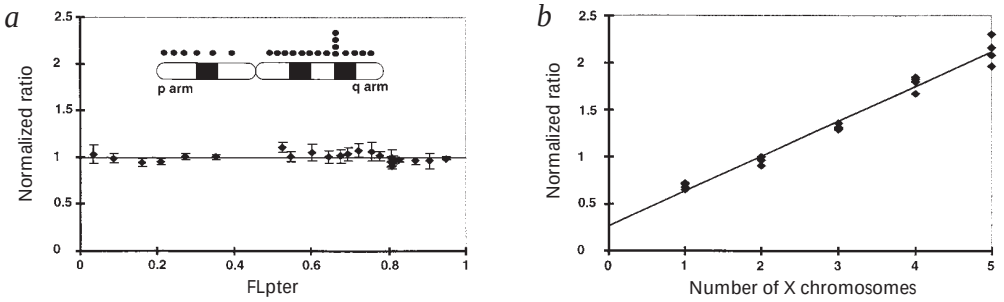

Fig. 2 Precision and accuracy of human copy number measurements. ***a***, Fluorescence ratios on the chromosome 20 array targets for six comparisons of a normal genome to itself. The inset schematically shows the location of the chromosome 20 array targets. The vertical set of targets on the q arm indicates the location of four nearly contiguous clones at 20q13.2. The ratios were normalized so that the average for all targets in each hybridization was 1.0. The data points show the mean of the six normalized ratios obtained for each target, the error bars indicate the standard deviations. They are plotted at the physical location of the clone as determined by FISH, measured as a fraction of the chromosome length relative to the p terminus (Flpter). The solid horizontal line is drawn at ratio 1.0. ***b***, Normalized fluorescence ratios on X chromosome targets as a function of X chromosome copy number. Arrays containing four clones from the X chromosome, and four clones on chromosome 20p (RMC20P107, RMC20P160, RMC20P178 and RMC20P099; Table 1), were hybridized with test genomic DNA from a normal male, normal female and three cell lines containing three, four and five copies of the X chromosome. Normal female reference DNA was used for all hybridizations. The fluorescence ratios on each of the X chromosome targets was normalized by the mean ratio of the chromosome 20 targets in that hybridization. All measurements are plotted, in some cases overlapping points obscure each other. The line is a linear regression through all of the data, slope 0.37, intercept 0.27.

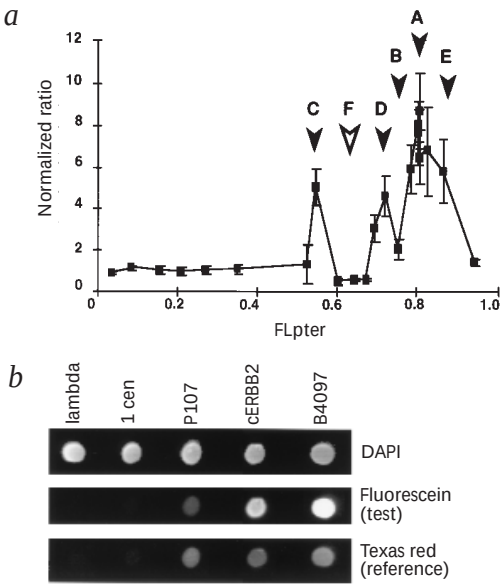

Fig. 3 Analysis of breast cancer cell line BT474. ***a***, Copy number variation on chromosome 20. The arrows A–E show regions of copy number increase[3], whereas F shows the location of copy number decrease. The data points indicate the means of three hybridizations and the error bars show the standard deviations. The points are connected by straight lines that do not provide any information about the copy number of sequences between the measured loci. The ratios were normalized to the average of the ratios of the six targets on chromosome 20p. ***b***, Montage of images of five array targets from an analysis of breast cancer cell line BT474. The targets are λ DNA, DNA from plasmid PUC 1.77, which contains the abundant repetitive sequence just below the centromere in chromosome 1, P1 RMC20P107 on chromosome 20p, cosmid cRC-Neu1, which contains the oncogene cERBB2 and BAC RMC20B4097, which is located at the peak of region A (Fig. 3*a*). The upper band shows the DAPI image of each target. (DAPI stains DNA in general.) The middle and lower bands show the corresponding fluorescein and Texas red signals from the BT474 and normal human female reference DNA, respectively. The signals on RMC20P107 are representative signals for targets at average copy number in the genome. In comparison, the λ target has a barely perceptible signal, less than 10% of the single copy level, from the human probes. Similarly, the signal on the chromosome 1 centromeric repetitive target, PUC1.77, has been suppressed to below 10% of the single copy intensity level. In contrast, the fluorescein signal on the cERBB2 target is significantly elevated above single copy level, whereas the Texas red reference signal is not, indicating the well-known amplification of cERBB2 in this cell line. Finally, the fluorescein signal on RMC20B4097 is above saturation, whereas the reference signal is comparable to the other human targets (its somewhat increased intensity is due primarily to its longer length).

aberrations on chromosome 20. Several distinct peaks and a copy number decrease were seen after analysis of breast cancer cell line BT474 (Fig. 3*a*). Labels A–E indicate locations where copy number increases in breast cancer cell lines and tumours had previously been found[3]. The deletion F, which was confirmed by interphase FISH using clones RMC20P107 (on 20p) and RMC20P154 (in region F), represents a new feature not previously recognized as significant in breast cancer. This region is known to be recurrently deleted in myeloid leukaemia[11,12]. FISH data[2] for chromosome 20 copy number in this cell line are in good agreement with the array data (Fig. 3*a*), further validating our procedures. Amplifications of cERBB2 and clone RMC20B4097 relative to clone RMC20P107 in BT474 are easily detected by visual inspection of the hybridization images (Fig. 3*b*).

Analysis of a set of four breast tumours, all selected by FISH using clone RMC20B4097 to have an amplification in region A, is shown (Fig. 4). Two regions of elevated copy number, D and E, had been discounted in previous work as loci that probably contain important oncogenes, because high-level amplification was not detected in tumours. In contrast, this study found both could be highly amplified (more than threefold, relative to 20p). Tumour S59 has an indication of the deletion F. Increases and decreases in copy number on chromosome 20 in this set of tumours are summarized (Table 1).

Data from nearly contiguous targets[9] in region A show the capability of array CGH for high resolution analysis of the boundaries and the amplitude structure of amplicons. In tumour S21, elevated copy number was found only for RMC20B4097 and RMC20B4130, indicating that the proximal amplicon boundary is near the neighbouring ends of RMC20B4097 and RMC20P4039, which are separated by less than 100 kb (ref. 9). Similarly, in S59, elevated copy number was found for the three distal clones but not RMC20B4135. Tumour S50, on the other hand, was elevated in copy number on all four clones in region A, but the copy number varied twofold (Fig. 4, inset).

We have developed an array CGH procedure that provides high resolution (approximately 40 kb) measurement of gains and losses of DNA sequence in genomes of mammalian complexity. Fluorescence ratios were linearly proportional to copy number over a dynamic range of several orders of magnitude, but the

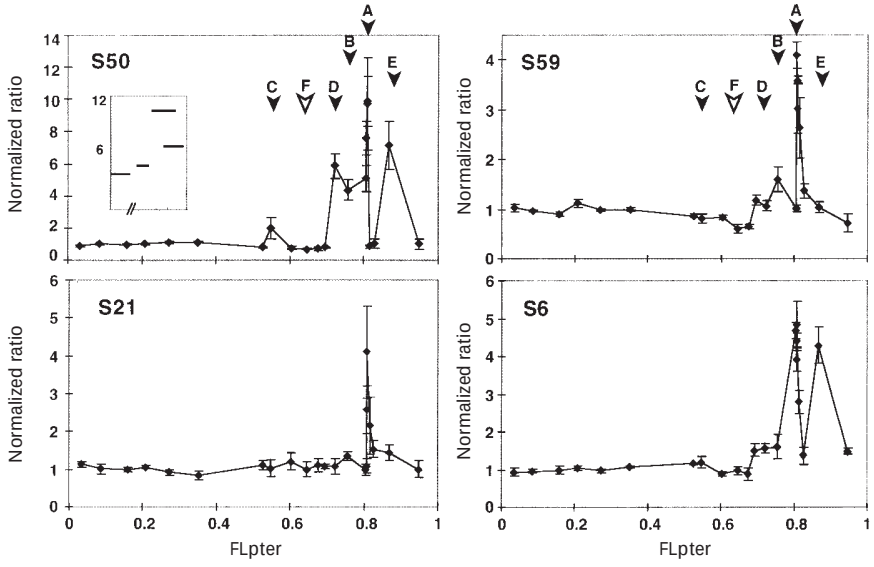

Fig. 4 Copy number variation on chromosome 20 in four breast tumours. Regions A–F are labelled as in Fig. 2. All specimens were selected by FISH for amplification in Region A. Array measurements, normalized to the average of the six targets on the 20p, reveal other well-defined regions of copy number change. The data points show the means and the error bars show the standard deviations of three independent measurements of each tumour. The inset for S50 shows an expanded view of the copy number changes for the four targets in region A. The horizontal bars indicate the approximate length and relative locations of the targets, the heights show the measured ratios.

slope of the relationship was somewhat reduced from ideal in the human genome. Single-copy decreases and increases from diploid were reliably detected. Within current measurement capability, all of the approximately 100 human genomic cosmid, P1 and BAC clones that we have used for targets have provided useful information. More work is needed, however, to determine if subtle clone-specific hybridization variation, due perhaps to a clone's exact repertoire of repetitive sequences, average base composition and so on, will impact the ability to consistently achieve the high measurement precision we have demonstrated.

Array CGH measurements of one cell line and a small set of breast tumours has provided new information on copy number aberrations involving chromosome 20. Most specifically, regions D, E and F were highlighted. More generally, the array CGH profiles of chromosome 20 (Fig. 4, Table 1) show the complexity of the copy number structure that can occur in this small region of the genome. These observations, together with the previous data obtained by FISH (refs 2,3), suggest that if copy number changes occur on chromosome 20 in breast cancer, multiple regions may be involved. Moreover, several frequently amplified candidate oncogenes not contained in the regions identified in our study also map to chromosome 20q. These include AIB1 (ref. 13) at 20q12, and CAS (ref. 14), TFAP2C (ref. 15) and BTAK (ref. 16), all at 20q13. Thus, a higher resolution array will probably reveal even more complex events than those shown (Fig. 3 and Fig. 4). Copy number increases on 20q have been associated with high S phase fraction, high grade and early recurrence in node negative breast cancer patients[17], with aggressive tumour behaviour[18], and with immortalization or extension of life span of primary epithelial cells in culture[19–21]. Thus, improved understanding of these aberrations might provide significant new information on genes that may contribute to one or more aspects of these phenotypes. Aberrations in other regions of the genome have also revealed significant complexity on close examination[4,22–24]. Therefore, the number of genes worthy of identification, and the complexity of the genetic aberrations that need to be considered in genotype/phenotype correlational studies, may be significantly greater than previously recognized. Data from region A in tumours S21, S50 and S59 show that array CGH can provide high-resolution, quantitative information to assist in understanding the biological and clinical consequences of these abnormalities and to refine the locations of the critical genes they contain.

Although the emphasis of this publication has been on the analysis of cancer, the capabilities of array CGH make it ideal for clinical genetic applications such as identification of disease genes, association of genetic aberrations with phenotype and analysis of cytogenetically cryptic copy number changes, for example, those that may occur near chromosome telomeres or translocation breakpoints. Thus, important applications in multiple areas suggest the value of implementing genome-wide array CGH.

Methods

Target clones. cERBB2, cosmid cRCNeu1 was obtained from R. White. X chromosome, one clone for the BTK locus at Xq21–23 (RMCXP001) was selected from the DuPont B P1 library (coordinates 58H8); three non-overlapping BAC clones, 590B9, 602N22 and 689C14 from library BC, for the Oculocerebrorenal syndrome locus at Xq26.1. Chromosome 20, see Table 1.

Specimens. λ DNA 15612-013 was purchased from GIBCO-BRL. Breast cancer cell line BT474 was grown using standard techniques. Normal human male and female DNA was isolated from lymphocytes obtained from a blood bank. Cell lines GMO4626, GMO1415E and GMO5009C, which have 3, 4 and 5 copies, respectively, of the X chromosome in a diploid background, were obtained from NIGMS Human Genetic Mutant Cell Repository, Coriell Institute for Medical Research. Genomic DNA was isolated from cells and cell lines using standard techniques. Snap-frozen blocks of primary breast tumour tissue from tumours with a copy number increase in region A (determined using FISH with clone RMC20B4097) were cryosectioned (4 µ) and stained with haematoxylin and eosin to determine the distribution of malignant to non-malignant cells in the section. Blocks were trimmed to remove areas containing normal ducts and lobules, and 10–20 sections of the trimmed block were collected in a microfuge tube for DNA extraction. Genomic DNA was isolated using QUIamp tissue kits (29304, Qiagen), including incubation with RNAase A, following the instructions of the manufacturer, except that lysis times were sometimes extended depending on the condition of the tissue. DNA was eluted from the column in water after 5 min incubation at 70 °C. The elution step was repeated to increase yield.

Array fabrication. Cloned genomic DNA for targets was isolated from bacterial cultures (500 ml) using Qiagen maxi kits (12162) following the instructions of the manufacturer, except that the volume of lysis buffer was increased 1.5–2-fold. The DNA was resuspended in TE buffer (400 µl, 10 mM Tris, 1 mM EDTA, pH 8), extracted with phenol, chloroform and isoamyl alcohol (25:24:1) and precipitated with ethanol. After resuspension in TE buffer, the DNA concentration was determined using a Hoefer TKO 100 fluorometer. This procedure typically yielded 40–80 µg of P1 or BAC DNA of sufficient purity to produce targets that bound effectively to the slides, and which had acceptably low autofluorescence. Target solutions were made by precipitating DNA (10 µg) in ethanol and dissolving the pellet in water (1 µl), followed by addition of DMSO (4 µl) containing nitrocellulose (approximately 0.4 µg/µl). The nitrocellulose solution was prepared by dissolving a nitrocellulose membrane (Gibco BRL 41051-012) in DMSO. Target solutions were stable and were used over periods over 1 yr without degradation in performance. Glass, quartz and fused silica slides were cleaned by incubation in 50% concentrated sulfuric acid/50% hydrogen peroxide overnight, washed 10 times in distilled water, air dried and heated to 90 °C for 15 min. The surfaces of the cleaned slides were coated by immersion in 95% acetone with 0.1% aminopropyltrimethoxy silane for 2 min at RT, after which they were washed five times in acetone and air dried. Quadruplicate spots (200–400 µm diameter) of each target solution were made by depositing the target solutions onto slides using capillary tubes. After air drying, the target spots were primarily DNA, containing less than 15% nitrocellulose by mass. They were invisible on the slide surface. As compared with other methods[6,7], the nitrocellulose substantially improved the ability to retain hybridizable DNA fragments in the targets. Signal intensities covering a dynamic range of at least 10^4 could be achieved by varying probe concentration. However, fragments smaller than approximately 2 kb were not bound effectively, limiting attachment of small PCR products, oligonucleotides and so on. No additional denaturation of target DNA was required before hybridization.

Comparative hybridization. Test and reference genomic DNA were labelled by nick translation with fluorescein dCTP (DuPont NEN NEL424) and Texas red dCTP (DuPont NEN NEL426), respectively. Unincorporated nucleotides were removed using a Sephadex G-50 spin column. Labelled DNA (200–400 ng) was mixed with Cot-1 DNA (35–70 µg; Gibco BRL) and precipitated with ethanol. (The amount of Cot-1 DNA was based on fluorometric determinations in our laboratory. We recommend checking each lot of Cot-1, because use of a sufficient amount is critical to obtain adequate suppression of repetitive sequences.) The precipitated DNA was dissolved in hybridization mix (10 µl) to achieve a final composition of 50% formamide, 10% dextran sulfate, 2×SSC, 2% SDS and 100 µg tRNA. The hybridization solution was heated to 70 °C for 5 min to denature the DNA, then incubated at 37 °C for approximately 1–5 h to allow blocking of the repetitive sequences. A wall enclosing the array (approximately 0.5 cm^2) was made with rubber cement, and the resulting well filled with hybridization mix. No coverslip was used. The slide was placed in a small sealed slide box containing 50% formamide (200 µl) and 2×SSC to prevent evaporation. Hybridization proceeded at 37 °C for 16–72 h on a slowly rocking table (2–4 cycles/min) to actively transport the hybridization mix over the array[25]. This resulted in a several-fold increase in signal intensity. After hybridization, the slide was washed once in 50% formamide, 2×SSC, pH 7, at 45 °C for 15–30 min, and once in 0.1 M sodium phosphate buffer with 0.1% NP40, pH 8 at room temperature for 5–10 min. Excess liquid was drained from the slides and the array mounted in an antifade solution[26] containing DAPI (1 µg/ml) to counterstain the DNA targets. A glass coverslip was sealed in place with nail polish. If necessary, the surface of the

cover slip and back surface of the slide were cleaned with lens cleaner (D53,881 Edmund Scientific) to remove fluorescent debris and films. Mounted slides were stable for weeks.

Imaging and analysis. Twelve bit fluorescence intensity data were obtained using a CCD camera (Sensys, Photometrics, equipped with a Kodak KAF 1400 chip) coupled to a 0.5× or 1× magnification optical system[27] (0.5× was obtained using an Ercona 25 mm focal length, f/1 C mount video lens combined with a Cannon 50 mm focal length, f/1.2 photographic lens; 1× was obtained using a pair of 25-mm focal length video lenses). Center to center pixel spacing referred to the array was approximately 7 µm in at 1× magnification, approximately 14 µm at 0.5×. Data from an area up to 14 mm×14 mm could be imaged in a single exposure at 0.5× magnification. A Chroma Technology P8100 multiband DAPI/Fluorescein/Texas red combined with a 625-nm short pass filter was mounted between the lenses to block excitation light. The use of the 625-nm filter substantially reduced interference from red autofluorescence originating in the glass slides and coverslips, permitting their use in place of quartz or fused silica. Excitation light was supplied by a HBO 100 mercury arc lamp. Kohler illumination optics provided uniform excitation intensity over the entire imaged area. The back side of the slide was coupled to the hypotenuse of a right angle fused silica prism with index matching oil (Cargille 06350 fused silica matching liquid). Excitation light entered one surface of the prism, passed through the slide and the array, underwent total internal reflection at the outer surface of the cover slip and passed back through the array. Images of the three fluorochromes were collected sequentially using a computer controlled filter wheel to select the appropriate Chroma Technology 8100 series excitation filter for DAPI, fluorescein or Texas red. Signals from test and reference DNA (fluorescein and Texas red, respectively) at the single copy level were detectable with exposures as short as 0.01 s. However, quantitative data were acquired with exposure times of 0.3–2.0 s. Images were analysed with custom software that segmented the array targets based on the DAPI staining, estimated and subtracted the background in the fluorescein and Texas red images, and calculated the total intensity and the intensity ratio of fluorescein and Texas red for each target. In addition, the Pearson's 'r' correlation[28] of a scatter plot of the fluorescein *versus* Texas red signal intensities for the pixels in each target was calculated. Data from targets with 'r' values below 0.8 were discarded, because this value indicated an unreliable measurement, usually due to fluorescent debris. For each clone, ratios on replicate targets (typically four) with $r > 0.8$ were averaged. Greater that 95% of target spots on an array met this criterion.

Acknowledgements

Work supported by NIH grants HD 17665, CA 45919 and P50 CA 58207; by U.S. DOE DE-AC03-76SF00098; California BCRP grants 1IB-003 and 2RB-0225; NIST ATP 94-05-0021; and Vysis. S.C. was supported by a postdoctoral fellowship from the U.S Army DAMD 17-96-1-6165 and Y.Z. received postdoctoral support from NIH training grant CA 09215. We thank U.J. Kim for the Xq26.1 clones.

Received 21 July; accepted 17 August, 1998.

1. Kallioniemi, A. *et al*. Comparative genomic hybridization for cytogenetic analysis of solid tumors. *Science* **258**, 818–821 (1992).
2. Tanner, M.M. *et al*. Increased copy number at 20q13 in breast cancer: defining the critical region and exclusion of candidate genes. *Cancer Res*. **54**, 4257–4260 (1994).
3. Tanner, M.M. *et al*. Independent amplification and frequent co-amplification of three nonsyntenic regions on the long arm of chromosome 20 in human breast cancer. *Cancer Res*. **56**, 3441–3445 (1996).
4. Bärlund, M. *et al*. Increased copy number at 17q22–q24 by CGH in breast cancer is due to high-level amplification of two separate regions. *Genes Chromosomes Cancer* **20**, 372–376 (1997).
5. Ning, Y. *et al*. A complete set of human telomeric probes and their clinical application. *Nature Genet*. **14**, 86–89 (1996).
6. Solinas-Toldo, S. *et al*. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**, 399–407 (1997).
7. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
8. Lockhart, D.J. *et al*. Expression monitoring by hybridization to high density oligonucleotide arrays. *Nature Biotechnol*. **14**, 1675–1680 (1996).
9. Collins, C. *et al*. Positional cloning of ZNF217 and NABC1: Genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc. Natl Acad. Sci. USA* **95**, 8703–8708 (1998).
10. Mohapatra, G. *et al*. Analysis of brain tumor cell lines confers a simple model of relationships among fluorescence *in situ* hybridization, DNA index, and comparative genomic hybridization. *Genes Chromosomes Cancer* **20**, 311–319 (1997).
11. Bench, A.J. *et al*. A detailed physical and transcriptional map of the region of chromosome 20 that is deleted in myeloproliferative disorders and refinement of the common deleted region. *Genomics* **49**, 351–362 (1998).
12. Wang, P.W. *et al*. Refinement of the commonly deleted segment in myeloid leukemia with a del(20q). *Genes Chromosomes Cancer* **21**, 75–81 (1998).
13. Anzick, S.L. *et al*. AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science* **277**, 965–968 (1997).
14. Brinkmann, U., Gallo, M., Polymeropoulos, M.H. & Pastan, I. The human *CAS* (cellular apoptosis susceptibility) gene mapping on chromosome 20q13 is amplified in BT474 breast cancer cells and part of aberrant chromosomes in breast and colon cancer cell lines. *Genome Res*. **6**, 187–194 (1996).
15. Williamson, J.A. *et al*. Chromosomal mapping of the human and mouse homologues of two new members of the AP-2 family of transcription factors. *Genomics* **35**, 262–264 (1996).
16. Sen, S., Zhou, H. & White, R.A. A putative serine/threonine kinase encoding gene *BTAK* on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. *Oncogene* **14**, 2195–2200 (1997).
17. Tanner, M. *et al*. Amplification of chromosomal region 20q13 in invasive breast cancer: prognostic implications. *Clin. Cancer Res*. **1**, 1455–1461 (1995).
18. Courjal, F. *et al*. DNA amplifications at 20q13 and MDM2 define distinct subsets of evolved breast and ovarian tumours. *Br. J. Cancer* **74**,1984–1989 (1996).
19. Reznikoff, C.A. *et al*. Long-term genome stability and minimal genotypic and phenotypic alterations in HPV16 E7-, but not E6-, immortalized human uroepithelial cells. *Genes Dev*. **8**, 2227–2240 (1994).
20. Savelieva, E. *et al*. 20q gain associates with immortalization: 20q13.2 amplification correlates with genome instability in human papillomavirus 16 E7 transformed human uroepithelial cells. *Oncogene* **14**, 551–560 (1997).
21. Salinas-Toldo, S., Dürst, M. & Lichter, P. Specific chromosomal imbalances in human papillomavirus-transfected cells during progression toward immortality. *Proc. Natl Acad. Sci. USA* **94**, 3854–3859 (1997).
22. Bièche, I. *et al*. Two distinct regions at 17q11–q12 involved in human primary breast cancer. *Cancer Res*. **56**, 3886–3890 (1996).
23. Karlseder, J. *et al*. Patterns of DNA amplification at band q13 of chromosome 11 in human breast cancer. *Genes Chromosomes Cancer* **9**, 42–48 (1994).
24. Reifenberger, G. *et al*. Refined mapping of 12q13-15 amplicons in human malignant gliomas suggests *CDK/SAS* and *MDM2* as independent amplification targets. *Cancer Res*. **56**, 5141–5145 (1996).
25. Albertson, D.G., Fishpool, R.M. & Birchall, P.S. Fluorescent *in situ* hybridization for the detection of DNA and RNA. in *C. elegans: Modern Biological Analysis of an Organism* (eds Epstein, H.F. & Shakes D.C.) 339–364 (Academic Press, New York,1995).
26. Johnson, G. *et al*. A simple method of reducing the fading of immunofluorescence during microscopy. *J. Immunol. Methods* **43**, 349–350 (1981).
27. Wittrup, K.D., Westerman, R.J. & Desai, R. Fluorescence array detector for large-field quantitative fluorescence cytometry. *Cytometry* **16**, 206–213 (1994).
28. Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. *Numerical Recipes in C: the Art of Scientific Computing 2nd edition* 636–639 (Cambridge University Press, Cambridge, 1994).
29. Shepherd, N.S. *et al*. Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc. Natl Acad. Sci. USA* **91**, 2629–2633 (1994).
30. Kim, U.J. *et al*. A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc. Natl Acad. Sci. USA* **93**, 6297–6301 (1996).
31. Stokke, T. *et al*. A physical map of chromosome 20 established using fluorescence *in situ* hybridization and digital image analysis. *Genomics* **26**, 134–137 (1995).