Cytometry 19:10–26 (1995)

Computer Image Analysis of Comparative Genomic Hybridization

Jim Piper, Denis Rutovitz, Damir Sudar, Anne Kallioniemi, Olli-P. Kallioniemi, Frederic M. Waldman, Joe W. Gray, and Dan Pinkel

Division of Molecular Cytometry, Department of Laboratory Medicine, MCB-230, University of California San Francisco, San Francisco, California 94143-0808 (D.S., A.K., O-.P.K., F.M.W., J.W.G., D.P.); MRC Human Genetics Unit, Western General Hospital, Edinburgh EH4 2XU, Scotland, U.K. (J.P., D.R.); Lawrence Berkeley Laboratory, Berkeley, California 94720 (J.P., D.S., J.W.G., D.P.); Department of Biomedical Sciences, University of Tampere, P.O. Box 607, SF-33101 Tampere, Finland (A.K.); Cancer Genetics Laboratory, Department of Laboratory Medicine, Tampere University Hospital, SF-33521 Tampere, Finland (O-.P.K.)

Received for publication March 21, 1994; accepted June 4, 1994

We describe and evaluate the image-processing and analysis techniques we have developed for the quantitative analysis of comparative genomic hybridization (CGH; Science 258:818, 1992). In a typical CGH application, two genomic DNA samples are simultaneously hybridized to metaphase chromosomes and detected with different fluorochromes. The primary data in CGH are contained in the intensity ratios of the fluorochromes as a function of position on the chromosomes, which reflect variation in DNA copy number ratio between the two DNA samples. Analysis involves chromosome segmentation, intensity normalization, background corrections, and calculation of the fluorescence intensity profiles and the ratio profile along the chromosome's length. Profiles from several copies of the same chromosome in different metaphases are averaged to reduce the noise. Confidence intervals are calculated and displayed for the mean profiles. The techniques were evaluated by examining the variability found in comparisons of two normal genomic DNAs, where the ratio was expected to be constant, and by measuring the ratios obtained for cell lines with cytogenetically documented copy number changes involving several chromosomal segments. The limits of sensitivity of CGH analysis were investigated by simulation. Guidelines for the interpretation of CGH data and indications of areas for future development of the analytical techniques are also presented. © 1995 Wiley-Liss, Inc.

Key terms: Amplification, background correction, deletion, fluorescence in situ hybridization, genomic DNA, image analysis, molecular cytogenetics, intensity profile, ratio profile, simulation, tumor cytogenetics

Since its recent introduction (4), comparative genomic hybridization (CGH) has become an important technique for genetic analysis. CGH provides a rapid method for comparing DNA sequence copy number throughout two (or more) genomes. It has found major application in the analysis of genetic aberrations in cancer (3–6,9). Detection and chromosome localization of copy number changes such as deletions and amplifications may highlight the locations of inactivated tumor suppressor genes and activated oncogenes, respectively. Because CGH does not require cell culture, it can be applied to many situations where standard cytogenetics yields little or no information, such as in solid tumors and archival (fixed) tumor specimens. Comparison of tumor and normal genomic DNAs is a major application of CGH, and, in this paper, we focus on the analysis of this type of experiment. However, CGH is also expected to have significant utility in perinatal analysis of whole and segmental chromosome aneuploidies such as Down syndrome and possibly also in microdeletion syndromes.

In a typical CGH analysis, an abnormal test genome and a normal reference genome are simultaneously hybridized to normal metaphase "target" chromosomes (Fig. 1) and detected with different fluorochromes. The ratio of the staining intensities at a particular location on a metaphase chromosome reflects the hybridization ratio between the two DNAs. This ratio is governed by the hybridization kinetics, which depend on the ratio of the copy numbers of the sequences that bind at that location

This work was performed under the auspices of the Office of Health and Environmental Research, U.S. Department of Energy, under contract DE-AC-03-76SF00098, with support from Imagenetics.

Address reprint requests to D. Pinkel, Division of Molecular Cytometry, MCB-230, UC San Francisco, San Francisco, CA 94143-0808.

in the two genomes. If normal genomic DNA, which has two copies of each unique sequence throughout the genome, is used as the reference DNA, then a change in the fluorescence intensity ratio between two chromosome locations indicates that the test DNA sequence copy numbers differ at those locations.

If the copy number variation is large, such as a high level gene amplification in a tumor, it can be clearly seen as intensity changes of the signal due to the test DNA alone. However, lower level changes such as deletions or duplications may be lost in the signal variation caused, for example, by regional variations in accessibility of the target DNA to the probe. The use of normal DNA hybridization as a comparison gives a local reference that can compensate for these uninteresting variations. One important use of CGH analysis is the detection of fractional copy number gain or loss of DNA that results from clonal heterogeneity or the inclusion of normal cells in the primary tumor specimen. In these circumstances, the best possible resolution of relative copy number is required in order to detect copy number changes that occur in a subpopulation of the sampled cells.

Major changes in the relative intensities of the two fluorochromes can be seen directly in the microscope if the intensities are approximately in balance. However, the eye-brain combination is limited in its capabilities for detecting and quantifying color ratio changes, and even simple multicolor electronic imaging that allows contrast enhancement and adjustment of the intensity balance of the two fluorochromes is of substantial assistance for improving visual recognition of regions of copy number variation. However, measurement of intensity ratio has been shown to provide a more quantitative method of assessing sequence copy number ratio. Thus calculation of ratio profiles for each chromosome coupled with statistical analysis of the profiles has proven to be the most powerful way to detect and map DNA sequence copy number changes throughout the genome. Therefore, unlike most other cytogenetic assays in which visual analysis is adequate but automation is helpful, digital image-analysis techniques are essential for CGH.

All of the many steps involved in performing CGH are undergoing continuous development. Because the overall result depends on the performance of each step, optimization of the entire process requires an internally consistent and somewhat circular effort. Thus, accurate image analysis is required to assess the quantitative nature of the biochemical reactions, but testing of the analytical approaches requires making assumptions about the fidelity of the hybridizations. In spite of this problem, substantial and rapid progress has been made on all aspects of CGH analysis. The progress in the laboratory methods for CGH is reviewed elsewhere (7). In this paper, we describe our current approach to image-analysis aspects of CGH, testing its performance by analysis of cell lines with well characterized copy number changes. We discuss the approximations involved in the analysis, indicate profitable directions for future development, and explore the potential limits of CGH using model calculations.

MATERIALS AND METHODS

Specimen Preparation and In Situ Hybridization

Normal genomic DNA was isolated from human peripheral blood lymphocytes from a variety of subjects. Test genomic DNA was obtained for the experiments described here from the breast cancer cell line 600MPE and from three cell lines with karyotypes, respectively 45,X, 46,XX, and 47,XXX (Human Mutant Cell Repository, Camden, NJ). Cell line 600MPE is basically diploid, but cytogenetic analysis indicates four copies of chromosome 1q and deletions of 9p, 16q, and distal 11q. There may be other small abnormalities concealed in the marker chromosomes. The DNAs were labeled with biotin-14-dATP (Life Technologies Inc.), digoxigenin-11-dUTP (Boehringer Mannheim), FITC-dUTP, or Texas red-dUTP (both DuPont) by nick translation as described elsewhere (7). Sixty to two hundred nanograms of labeled test and reference DNA and 5–10 μg of unlabeled Cot-1 DNA were mixed, ethanol precipitated, dissolved in hybridization mix (70% formamide/2× SSC/10% dextran sulfate), and hybridized for 2 days to normal human metaphases on slides. After washing, slides employing biotin- and digoxigenin-labeled nucleotides were fluorescently stained with avidin-FITC (green) and antidigoxigenin rhodamine (red). The DNA was counterstained with DAPI, which produced bands adequate for chromosome identification.

Digital Imaging and Analysis

Overview. High-resolution digital images of the DNA counterstain and the fluorochromes in the test and reference DNA were acquired; these will be referred to as the *counterstain image*, the *test image*, and the *reference image*. The hybridization signals appeared as tiny fluorescent granules on the chromosomes; the better the hybridization quality, the denser the granules. The intensity ratio for each pixel tended to be highly variable given the granularity of signal, so it proved advantageous to perform two levels of signal averaging. In the first level, the chromosomes were treated as one-dimensional objects, and fluorescence intensity profiles were generated by integrating the pixel fluorescence intensities across the chromosome width. The second level took advantage of the fact that each copy of a particular chromosome, in fact each chromatid, should contain identical information no matter which metaphase spread it is from on one slide or even if it is from an entirely different hybridization using the same pair of DNA samples. Thus averaging profiles from multiple homologous chromosomes should suppress the noise inherent in the hybridizations and thereby improve determination of the ratios. However, the absolute value of the fluorescence intensities and the ratios may vary widely among these different chromosomes due to variations in background signals within and among metaphases, hybridization conditions, exposure times, bleaching, etc. Effective profile averaging thus required background correction and intensity normalization.

Image acquisition. Images were obtained on an automated Zeiss Axioplan microscope (C. Zeiss, Thornbor-

ough, NY) controlled by the QUIPS software package [Sudar D: QUIPS, a quantitative image processing system for acquisition and analysis of multicolor fluorescence images (in preparation)]. The microscope was equipped with a ×63 1.25 na Plan Neofluar objective; a Ludl computer-controlled stage (Ludl Electronic Products, Hawthorne, NY); a 100 W mercury arc lamp; Chroma Technology "P1" multibandpass beam splitter and emission filter (Chroma Technology, Brattleborough, VT) and single-band excitation filters for each fluorochrome, which were mounted in a computer-controlled filter wheel to permit preferential excitation of individual fluorophores; a Quantex ICCD camera and video monitor (Quantex, San Jose, CA) to give a live image for initial selection of a suitable metaphase and for focusing; and a Photometrics cooled slow-scan "scientific" CCD camera (Photometrics, Tucson, AZ) or a Xillix MicroImager 1400 CCD camera (Xillix, Vancouver, British Columbia, Canada). Images comprised up to 1,340 × 1,138 pixels with spacing of 0.1 μm referred to the specimen, though usually only a subset of the field was required for a metaphase (600 × 600 pixels was typical). Registration error between the fluorochrome-specific images is governed by the quality of the chromatic corrections in the microscope optics and has been measured to be less than 0.1 μm across the area of a metaphase (2). For the purposes described herein, this level of registration error was insignificant and was left uncorrected. Further characteristics of the imaging hardware will be described elsewhere (2).

Metaphases were located by visual scanning. Each metaphase was focused using the "live" ICCD image of counterstain fluorescence, and then a sequence of counterstain, test, and reference images was acquired under software control by selecting the appropriate excitation filters [Sudar D: QUIPS, a quantitative image processing system for acquisition and analysis of multicolor fluorescence images (in preparation)]. CCD camera integration times were determined automatically and were typically between 0.5 and 5 s (the higher times being required for directly labeled nucleotides). Each multicolor image was immediately filed to disk for subsequent analysis on a workstation not associated with a microscope. The digitization software has been used for a variety of applications that required high-quality, multicolor digital microscope images (8).

Contrast enhancement, image display, and chromosome identification. General-purpose software based on SCILimage [SCIL Image, DIFA Bveda, The Netherlands (15)] was used to display a pseudocolor composite image (Fig. 1) or any fluorochrome-specific image. Contrast was linearly stretched to fill the dynamic range of the computer display (Fig. 2). In particular, a contrast-enhanced image of the counterstain (DAPI) banding was used to identify the individual chromosomes in the metaphase. In normal operation, the counterstain image was displayed in blue, the test DNA image in green, and the reference DNA in red. Thus, in the composite image, those regions of the test genome that had higher copy number than average appeared to be greener, and those that were at lower copy number than average appeared to be redder than the orange hue typical of regions of average copy number (Fig. 1). The counterstain image was included to permit accurate visual identification of each chromosome. This was needed, because the test and reference images became very weak in regions with high concentrations of repetitive DNA such as centromeres (because the hybridization signal is blocked by the Cot-1 DNA), making recognition of associations between *p* and *q* arms ambiguous. The three-color display permitted rapid visual recognition of those regions of the test genome that had DNA sequence copy number that departed substantially from average, and it was also used to evaluate hybridization quality (7).

Image analysis. Quantitative analysis software was based on an existing library of chromosome analysis algorithms (12,13). The analysis consisted of the following steps.

Chromosome segmentation. Chromosomes were first segmented automatically using the "chromosome definition image" obtained by adding the counterstain and reference images after standardizing them so that they had the same range of pixel values. The two images were added, because experience showed that segmentation based on a DAPI counterstain image alone was sometimes inadequate due to the weak staining in some regions, particularly at the chromosome telomeres. However, the precise balance of the two images was not expected to have a major effect on segmentation, and adding the images after normalizing their respective gray scales was expected to be sufficient. It should be emphasized that adding the two images is not an essential part of the method, and, if an alternative counterstain adequately outlined the individual chromosomes, then that image alone could be used as the chromosome definition image.

Automatic segmentation was based on thresholding the pixel gray values of the chromosome definition image using a single global threshold for the entire image. We have previously found that choosing a threshold for a microscope image field on the basis of the gray value distribution alone tends to result in a threshold value that is too low; it is effectively representative of the distribution of the background portion of the image rather than the edges of the objects of interest and has to be corrected by addition of an offset. Automatic determination of a suitable offset has proved to be difficult. We therefore introduce here a new method for choosing the threshold. The gradient of the image at each pixel was calculated by a finite difference technique. A bivariate histogram of the number of pixels with given intensity and absolute value of gradient was formed (Fig. 3). A global intensity threshold was computed as the mean intensity value of that portion of the histogram in which the gradient exceeded the overall mean gradient (Fig. 3). Including gradient information in the choice of intensity threshold made it more likely that the chosen threshold value would be representative of the higher slope regions of the edges of the objects in the image.

The initial segmentation was followed by an adaptive

technique that used morphological opening and revised thresholding of apparent chromosome clusters in order to segment the metaphase more completely. Even after this automatic step, two or more chromosomes were frequently clustered into single objects. Remaining chromosome clusters that the operator deemed adequate for analysis were resolved interactively by using the mouse to draw suitable separation lines. Chromosomes with overlaps or with substantial lengths of contact with other chromosomes were not analyzed. Typically, more than 75% of the chromosomes in a metaphase were suitable for analysis, so that four or five profiles for each chromosome type could be obtained from three to four metaphase cells.

Normalization of hybridization intensities. The absolute fluorescence intensity of any of the fluorochromes in CGH is governed by illumination brightness, filter bandwidths, dye absorbance and quantum efficiency, camera sensitivity and exposure time, bleaching, hybridization efficiency, DNA labeling efficiency, and metaphase quality. Thus, absolute intensity alone does not carry any essential information. To compare information among different copies of the same chromosome type in different metaphases and in different experiments, some sort of normalization or standardization procedure is necessary. We chose a procedure that approximately equalized the median intensities of the test and reference images in all metaphases. Thus, the corrected median test/median reference ratio was approximately equal to 1.0 for each. Note, however, that this is not necessarily a copy number ratio of 1.0 between the DNA samples; determination of the DNA sequence copy number ratio that corresponds to a test/reference ratio of 1.0 requires additional information as described below.

Intensity normalization was accomplished in two stages. In the first, the segmentation mask obtained by thresholding the chromosome definition image as described in the previous section was applied to the test and reference images to form a mask containing the (incompletely segmented) metaphase chromosomes. Above threshold objects that did not belong to the target metaphase, for example, adjacent interphase nuclei, were excluded by operator interaction. In each of the test and reference images, the mean intensities outside the mask, B_t and B_r, were determined. The median signal intensities m_t and m_r of the images within the mask were then calculated as

$$m_t = \text{median } \{i_t\} - B_t,\ m_r = \text{median } \{i_r\} - B_r \qquad (1)$$

where $\{i_t\}$ and $\{i_r\}$ are, respectively, the sets of test and reference pixel intensities within the segmentation mask. The reciprocals of m_t and m_r were the intensity normalization factors. Note that they were obtained without the need to fully segment or analyze the entire metaphase.

Local background correction. The fluorescence background varied significantly over the metaphase field. The background level was particularly high relative to the signal when directly fluorochrome-conjugated nucleotides were used in CGH; sometimes it amounted to 40% of peak signal intensity (Fig. 2). We assumed that the background in the vicinity of each chromosome contributed additively to the measured pixel intensities within a chromosome. The local background, b, for a particular chromosome was defined to be the mode of the pixel values of the fall set boundary of the chromosome. The fall set of an image region (14) is the set of pixels that are reached by moving away from the region along paths of strictly decreasing intensity (Fig. 4). In the case of well isolated chromosomes, the fall set boundary provides a natural limit to the chromosome's "halo," the set of pixels surrounding the chromosome whose values are nevertheless a little above background. Here, the fall set boundary was defined in the chromosome definition image and replicated in the test and reference images (Fig. 5); it was usually considerably outside of the previously found segmentation boundary of the chromosome. The corrected test and reference intensity values i'_t and i'_r of a pixel in the chromosome were then calculated as:

$$i'_t = (i_t - b_t)/m_t,\ i'_r = (i_r - b_r)/m_r \qquad (2)$$

where i_t and i_r were the intensities of the pixel in the original test and reference images, b_t and b_r the local backgrounds, and m_t and m_r the median intensities of those images from equation 1.

The mode of the values on the fall set boundary was used for the local background, because it was expected that this would give the best value for those chromosomes that were not entirely isolated from others, as the higher values at the regions of contact would not affect the mode (Fig. 4). However, this will clearly be in error for chromosomes segmented from dense clumps where a substantial portion of the boundary is in contact with other chromosomes. Such chromosomes were not analyzed.

Profile extraction, ratio computation. The reference and test intensity profiles for each chromosome (Fig. 5e) were obtained by integrating the pixel values (corrected as described in equations 1 and 2) along slices orthogonal to the chromosome axis (Fig. 5b,c). The axis finding and pixel value interpolation methods used were as described previously by Piper and Granum (12). In short, the axis was derived from a skeleton of the shape of the segmented chromosome viewed as a binary image. Pixel values along the transverse slices were calculated by bilinear interpolation from the four surrounding pixels of the original digitization grid.

Defining the ends of chromosomes is critical for averaging profiles (see below) and association of ratio changes with genetic map location based on chromosomal fractional length. The chromosome end was defined to be at the points where the profile extracted from the chromosome definition image first reached 50% of the mean profile value.

Image display options for the analysis of individual chromosomes. The CGH analysis program was interactive. The cytogeneticist identified chromosomes based on the DAPI banding, assisted segmentation, selected chromosomes for profiling, and decided whether the ratio profile data could be kept for interpretation or

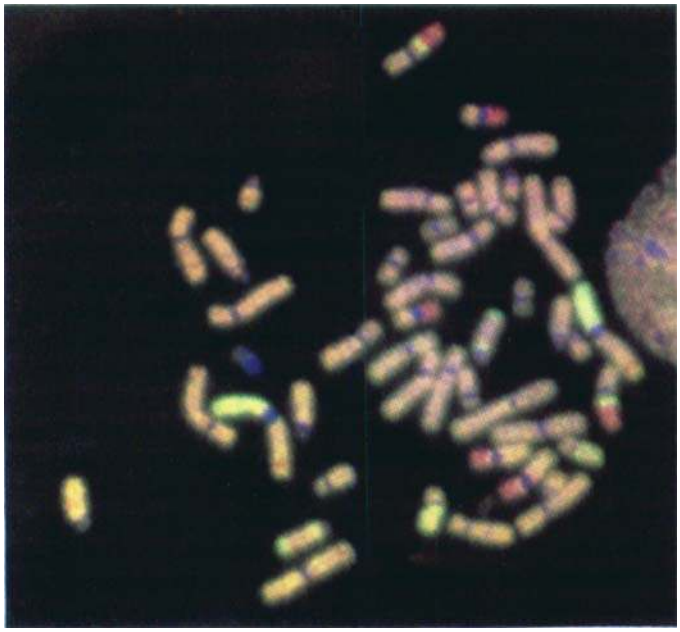

FIG. 1

FIG. 1. Three-color digital pseudocolor image of a double genomic hybridization using directly fluorochrome-conjugated nucleotides and DAPI counterstain. The test DNA hybridization is displayed in green, the reference in red, and the counterstain in blue. The test DNA came from breast cancer cell line 600MPE.

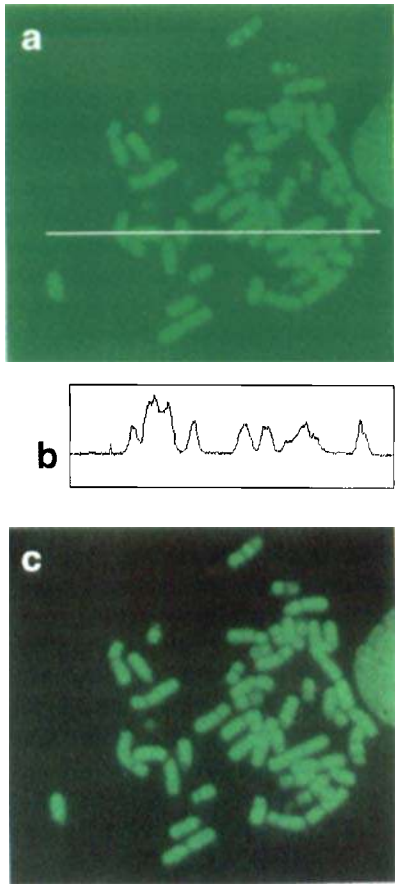

FIG. 2.

FIG. 2. **a:** The green test image of the metaphase shown in Figure 1, hybridized with FITC-dUTP directly labeled genomic DNA, showing a high background level. **b:** The relative intensities recorded along the line through the metaphase shown in (a). **c:** The contrast stretched version of a.

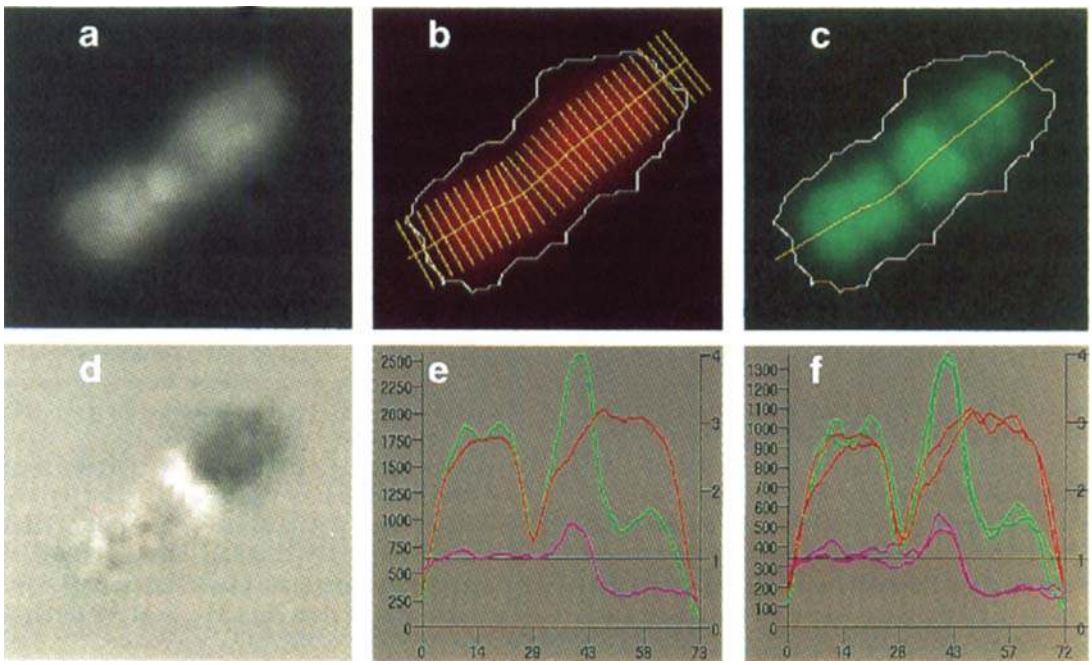

FIG. 5. Caption on facing page.

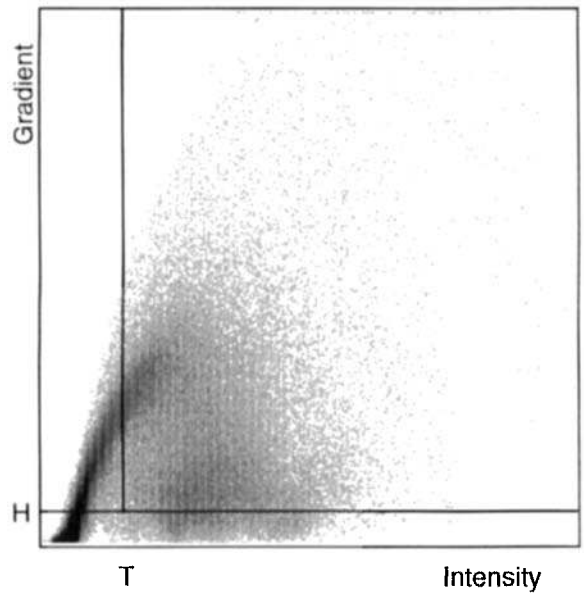

FIG. 3. Two-dimensional histogram of the number of pixels (represented by darkness value) that have a particular (intensity, gradient) value. H is the mean value of the gradient, and T the mean value of the intensity for those pixels whose gradient exceeds H. T is, thus, the chosen pixel value threshold. A fourth-root transformation of the darkness values has been used to make the wide variation of frequencies of (intensity, gradient) pixel values more apparent in this histogram; in actuality, almost all the pixels congregate in the bottom left corner.

FIG. 4. Boundary of fall set of a chromosome marks the natural limit of chromosome material.

should be rejected due to poor quality or artifact. Likely reasons for rejection included a chromosome being a member of a dense cluster (in which case, the local background estimation cannot be relied upon); poor fitting of the medial axis, resulting in unrepresentative profiles; or poor hybridization, resulting in an unacceptable level of "noise." For assistance with these judgments, a variety of display facilities was provided. These included enlarged images of the counterstained, test, and reference versions of the chromosome and the computed medial axis and transverse "slices" that were the basis of profile construction (Fig. 5a–c). The user could choose to display either the profiles from the whole chromosome (Fig. 5e) or, alternatively, the pairs of profiles obtained from each chromatid independently (Fig. 5f). A gray-scale representation of the difference between the normalized test and reference images of the analyzed chromosome assisted in interpretation of possible small regions of copy number change (Fig. 5d).

Profile averaging. The output from the interactive analysis program was a set of chromosome ratio profiles labeled by chromosome number. These could be displayed as a "copy number karyotype" (Fig. 6). The profile for each chromosome was normalized to a standard length proportional to the mean DNA content of the chromosome in the data published by Morton (10) with chromosome 1 arbitrarily assigned a length of 100. A normalized profile results from longitudinally stretching or compressing a profile to have the desired length with the new profile values computed from the old by interpolation.

There are clearly at least two ways of constructing an average ratio profile. Either the test and reference intensity profiles are first averaged and the mean ratio profile calculated from the mean intensity profiles, or the individual ratio profiles can be averaged. The first method might appear to be preferable, in that it would be less prone to the noise in the individual ratio profiles that results when both reference and test signals are small; taking a ratio of means is usually preferable to taking the mean of a set of ratios. However, the second method provides a straightforward way to estimate a confidence interval for the mean ratio at any part of the genome and was therefore chosen, because we believe that such confidence intervals are important for interpretation of the mean profiles. As is shown below, tests indicated little difference between the methods in practice.

The copy number karyotype program could display results from several metaphases simultaneously, either by superimposing the ratio profiles from a number of chromosomes of the same type (Fig. 7a) or by displaying an estimated mean ratio profile (Fig. 7b) together with a choice of variability measure at every point of the profile. Suppose that a set of n ratio profiles for chromosome c is represented as $[P_{jc}(i)]$, where i indexes the position along the chromosome and j indexes the profile in the set. The variability measurements included range, stan-

FIG. 5. Steps in the processing of a chromosome 11 from the metaphase in Figure 1. **a:** The DAPI image. **b:** The reference hybridization showing the chromosome axis, every third transverse slice from which profile data are accumulated, and the fall set boundary. **c:** The test hybridization. **d:** The difference image test − reference, with a difference of zero displayed in midgray. **e:** The test and reference whole chromosome profiles together with the ratio profile (colored lilac). **f:** The corresponding pairs of individual chromatid profiles.

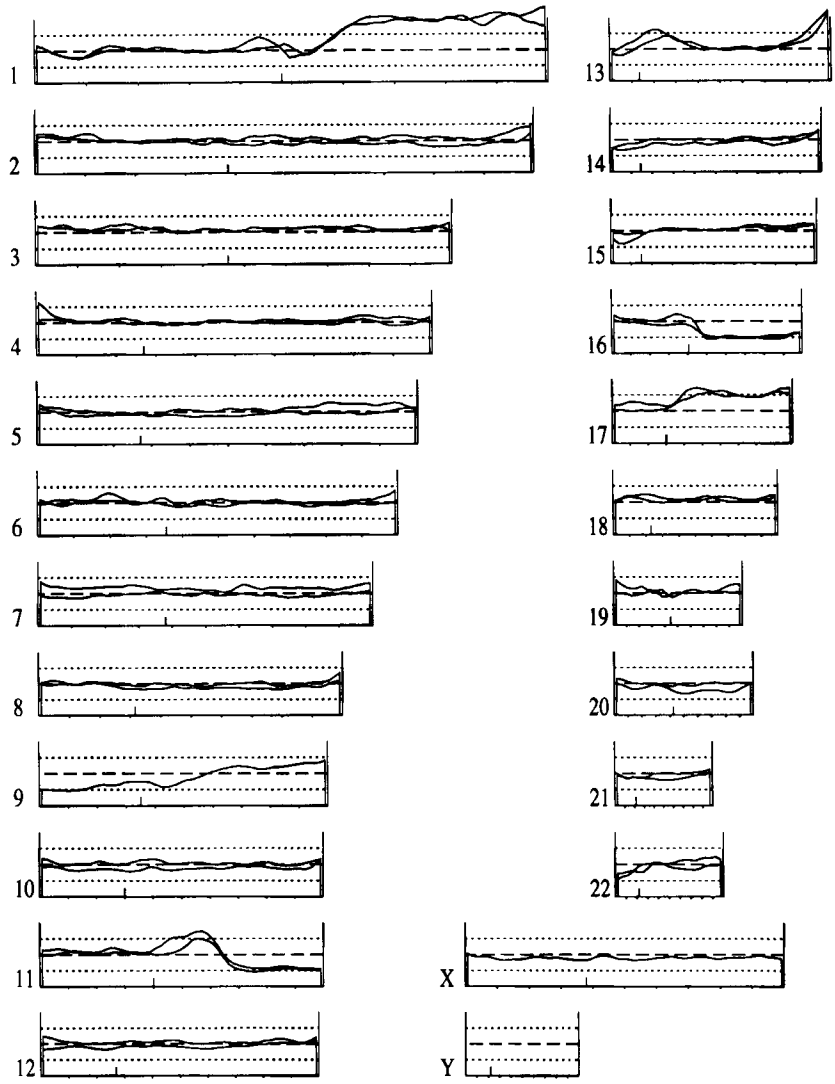

FIG. 6. Copy number karyotype of a single analyzed metaphase. Chromosome Y was not analyzed, because the test DNA was female and did not hybridize to the Y. The dashed lines indicate a normalized intensity ratio of 1.0, the dotted lines ratios of 0.5 and 1.5.

dard deviation, standard error of the mean, and 95% and 99% confidence intervals calculated along the profiles for each point i and chromosome c as follows:

$$\mathbf{P}_c(i) = \Sigma_j P_{jc}(i)/n$$

is the mean profile,

$$Max_c(i) = \max_j[P_{jc}(i)], Min_c(i) = \min_j[P_{jc}(i)]$$

are the range profiles,

$$\sigma_c(i) = \sqrt{\{\Sigma_j[P_{jc}(i) - \mathbf{P}_c(i)]^2/(n-1)\}}$$

is the standard deviation, and

$$[\mathbf{P}_c(i) \pm \sigma_c(i)]$$

are the standard deviation profiles,

$$\mathbf{s}_c(i) = \sigma_c(i)/\sqrt{n}$$

is the standard error of the mean, and

$$[\mathbf{P}_c(i) \pm \mathbf{s}_c(i)]$$

are the upper and lower standard error profiles,

$$[\mathbf{P}_c(i) \pm 2.\mathbf{s}_c(i)]$$

and

$$[\mathbf{P}_c(i) \pm 3.\mathbf{s}_c(i)]$$

are the upper and lower profiles for approximately 95% and 99% confidence limits of the estimated mean, respectively. These profiles, covering the entire genome

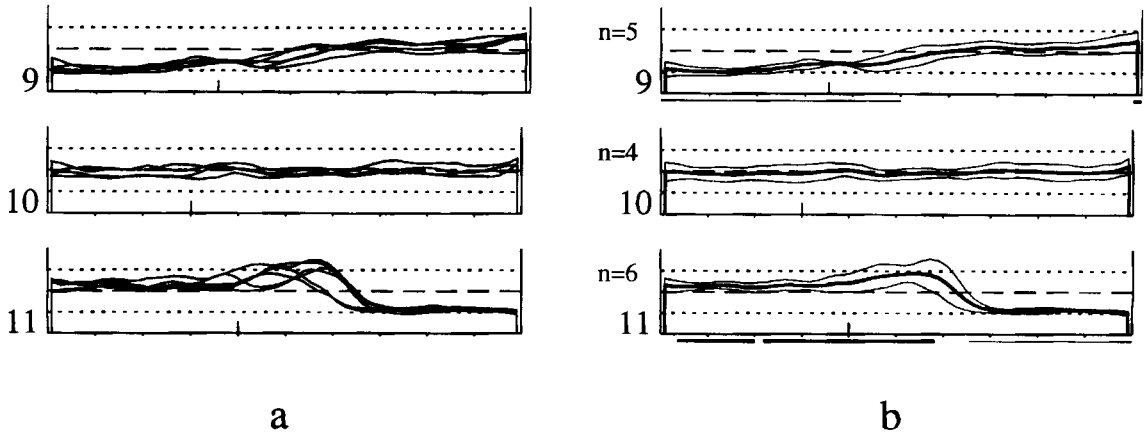

FIG. 7. Composite copy number karyotypes from several target metaphases used in a single CGH analysis. **a:** Multiple profiles. **b:** Mean profiles with 99% confidence limits. Underscoring shows regions whose confidence interval is either entirely less than 1.0 (thin) or more than 1.0 (thick). For reasons of space, only chromosomes 9–11 are shown.

from 1ptel to Yqtel, were the end result of the CGH analysis of DNA sequence copy number.

Variability of Ratio Measurements and the Effect of Profile Averaging

Averaging the ratio profiles from several homologous chromosomes accomplishes the dual goals of smoothing the local variations in ratio due to random noise sources and reducing the effect of circumstances that affect the whole profile—for example, consequences of the position of the chromosome in the microscope field such as illumination chromatic aberration (2) or consequences of the position in the metaphase such as local background variation or the effect of neighboring chromosomes; these we might regard as structured noise. Statistics characterizing the mean and variability of CGH ratios and the benefits of profile averaging were measured for each chromosome as follows (dropping the previous chromosome subscript c):

a) Let $[P_j(i)]$ be the set of n ratio profiles for a particular chromosome, length-normalized to common length l, where i indexes the position along the chromosome and j indexes the profile in the set. Then, $\mu_j = \Sigma_i P_j(i)/l$ is the mean of $P_j(i)$, $\mathbf{P}(i) = \Sigma_j P_j(i)/n$ is the mean profile (Fig. 7b), and $\boldsymbol{\mu} = \Sigma_i \mathbf{P}(i)/l$ is the overall mean of the set of ratio profiles for the chromosome.

b) By comparing the mean of the standard deviations of the individual profiles $\underline{\sigma} = 1/n\, \Sigma_j \sqrt{\{1/(l-1)\, \Sigma_i [P_j(i) - \mu_j]^2\}}$ with the standard deviation $\boldsymbol{\sigma} = \sqrt{\{1/(l-1)\, \Sigma_i [\mathbf{P}(i) - \boldsymbol{\mu}]^2\}}$ of the mean profile, we obtain an estimate of the degree of smoothing of the random fluctuations that occur along the length of each individual profile and the effects of profile averaging.

c) An indicator of positional or structured noise effects between different examples of the same chromosome is given by the standard deviation of the means of the individual profiles $\sigma' = \sqrt{[1/(n-1)\, \Sigma_j (\mu_j - \boldsymbol{\mu})^2]}$.

CGH analysis was performed in two experiments wherein both the test and reference DNA samples were derived from normal peripheral blood lymphocytes. The aim was 1) to determine how closely the overall mean ratio $\boldsymbol{\mu}$ was to 1.0 and if there were regions of the genome for which $\boldsymbol{\mu}$ differed from 1.0 consistently, and 2) to determine the noise inherent in CGH measurement and how this might be reduced by taking the average of ratio profiles, as measured by the variability statistics $\underline{\sigma}$, $\boldsymbol{\sigma}$, and σ'.

First, DNA from a single individual was compared. Whole genomic DNA was obtained from the lymphocytes of a single healthy female donor, split into several aliquots, and labeled with nucleotides directly conjugated with FITC (test samples) and Texas red (reference samples). Four metaphases were analyzed in each of two separate CGH preparations (Fig. 8, Table 1). Second, DNA from different normal individuals was compared. Whole genomic DNA was obtained from the lymphocytes of six healthy female donors. Five were labeled with biotin (test samples) and one with digoxigenin (the single reference sample). One metaphase was analyzed from each of the five hybridizations (Fig. 9, Table 2). In both cases, every nonoverlapping chromosome was identified and its ratio profile calculated.

Validation of Copy Number Estimation From Profiles

Our approach to quantitative CGH analysis rests on three assumptions; first, that unwanted background contributions to fluorescence signal intensity are additive to the signal and can be determined by examination of the intensity in the neighborhood of each chromosome; second, that resulting corrected signal intensity at a chromosomal location is directly proportional to DNA copy number of sequences in the hybridizing genome that bind there, modified only by factors that affect the binding of all hybridizing genomes equivalently; and, third, that our approximate intensity normalization procedure permits meaningful comparison of data between separate metaphases within and among experiments. If all of these as-

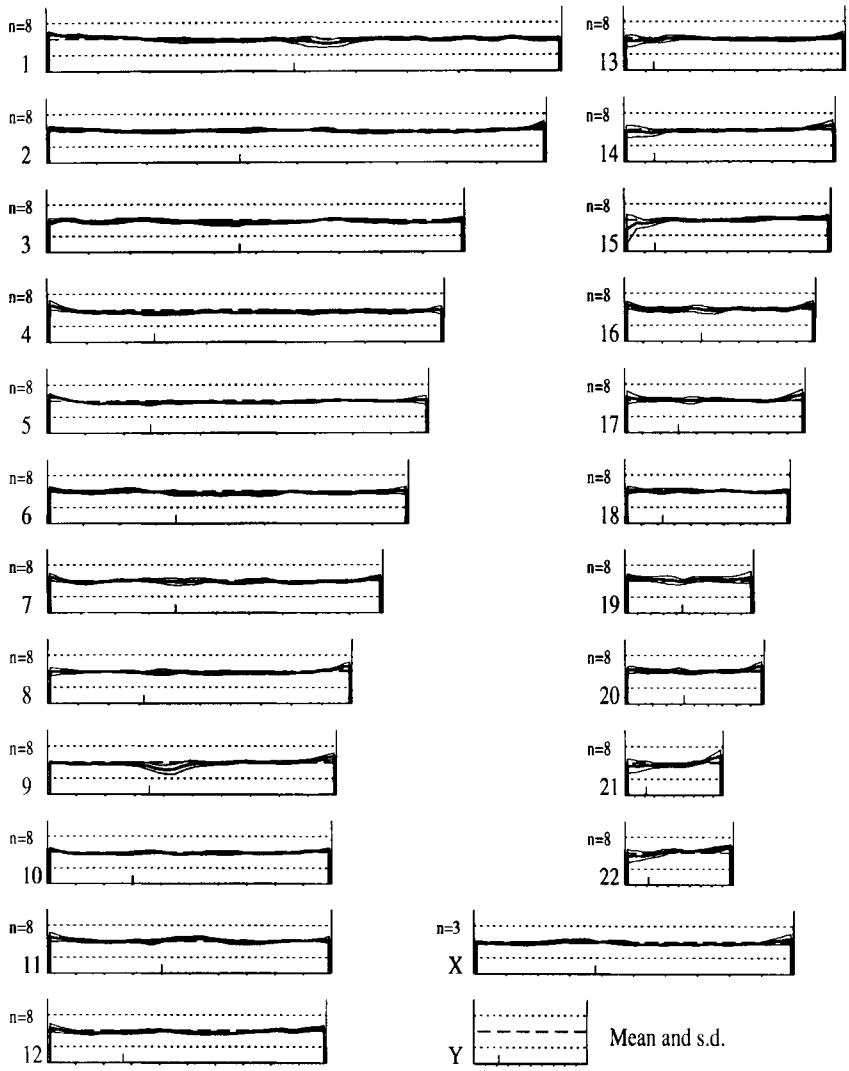

FIG. 8. Mean and standard deviation copy number karyotype profiles of normal female vs. normal female CGH analyses using directly labeled test and reference DNAs obtained from a single individual.

sumptions are satisfied, then the measured intensity ratio profile between a test and a normal genome using our approach should be directly proportional to the copy number profile of the test genome with the constant of proportionality the same for all measurements.

The degree to which actual preparation and analysis procedures meet the above criteria was determined by regression analysis of data comprising ratio measurements for test genomes with known regions of variant copy number. In the ideal case, when measuring a test genome of mean copy number equal to 2.0, the function relating measured ratio r to copy number n should be $r = 0.5n$ (because $r = 1$ implies two copies of a sequence when a normal genome is used as the reference). Test DNA samples were prepared using indirect fluorochrome labeling as described above from cell lines having, respectively, one, two, or three copies of the X chromosome (Human Mutant Cell Repository, Camden, NJ) and using direct fluorochrome conjugation from the cell line 600MPE, which has regions of one, two, and four copies of DNA sequences. The overall mean values of the ratio profile over the relevant chromosomal regions were obtained by averaging at least four chromosomes in every case. In the first set of hybridizations, Xp and Xq were treated separately. In the 600MPE cell line material, regions that comprised whole chromosome arms, or at least a large part of a chromosome long arm, were selected where cytogenetic analysis had confirmed that DNA was present in a single copy (9p, 16q, distal 11q), two copies (2q), or four copies (1q; see Figs. 1, 5–7).

Simulation Study of Potential Sensitivity

The lower limit of detectability of amplifications or deletions under optimal experimental conditions is not

Table 1
Overall Mean of Ratio Profiles, Together with Various Measures of Variability (see text), for Normal vs. Normal CGH Using Directly Labeled Test and Reference DNAs Obtained From a Single Individual[a]

		Short arm						Long arm					
Chromosome	N	μ	σ	$\bar{\sigma}$	σ'	μ^*	σ^*	μ	σ	$\bar{\sigma}$	σ'	μ^*	σ^*
1	8	1.01	0.06	0.07	0.02	1.02	0.06	1.00	0.02	0.04	0.03	0.98	0.05
2	8	0.99	0.02	0.05	0.02	1.00	0.03	0.98	0.03	0.05	0.01	0.98	0.02
3	8	0.97	0.04	0.06	0.05	0.99	0.04	0.96	0.04	0.05	0.03	0.96	0.03
4	8	0.95	0.05	0.06	0.03	0.96	0.05	0.94	0.02	0.04	0.03	0.93	0.02
5	8	0.96	0.05	0.06	0.02	0.97	0.06	0.97	0.03	0.05	0.01	0.96	0.03
6	8	1.00	0.03	0.06	0.03	1.00	0.02	0.95	0.03	0.05	0.03	0.94	0.03
7	8	0.99	0.03	0.06	0.03	0.99	0.03	0.98	0.03	0.05	0.03	0.97	0.02
8	8	0.98	0.01	0.04	0.02	0.98	0.01	0.97	0.04	0.06	0.03	0.96	0.02
9	8	0.96	0.01	0.03	0.03	0.96	0.01	1.02	0.05	0.06	0.03	0.98	0.05
10	8	0.96	0.02	0.04	0.02	0.97	0.03	0.99	0.03	0.05	0.02	0.98	0.02
11	8	1.01	0.03	0.05	0.03	1.00	0.04	1.03	0.05	0.06	0.02	1.03	0.04
12	8	0.95	0.03	0.04	0.05	0.94	0.03	0.98	0.04	0.06	0.02	0.96	0.02
13	8							0.96	0.03	0.05	0.03	0.96	0.02
14	8							0.98	0.04	0.05	0.02	0.96	0.01
15	8							1.01	0.03	0.05	0.03	1.00	0.03
16	8	1.00	0.02	0.04	0.07	1.00	0.03	1.02	0.03	0.04	0.04	1.00	0.02
17	8	1.03	0.01	0.04	0.05	1.03	0.01	1.01	0.03	0.06	0.04	1.01	0.01
18	8	1.01	0.01	0.04	0.06	1.02	0.02	1.00	0.02	0.05	0.03	1.00	0.01
19	8	1.04	0.04	0.06	0.08	1.06	0.02	1.05	0.03	0.05	0.08	1.03	0.03
20	8	1 02	0.01	0.06	0.03	1.03	0.01	1.02	0.03	0.04	0.05	1.00	0.02
21	8							0.98	0.07	0.08	0.04	0.93	0.01
22	8							1.08	0.04	0.05	0.05	1.05	0.05
X	3	0.98	0.03	0.04	0.03	0.98	0.04	0.96	0.03	0.05	0.02	0.95	0.03

[a]Because the DNA was obtained from a female, Y chromosome values have been omitted. Those chromosomal regions of substantial size where CGH analysis is not valid because Cot-1 blocking prevents hybridization by the genomic DNAs, i.e., the centromeric regions of chromosomes 1, 9, 16, and the acrocentric chromosome short arms, were excluded from the analysis.

yet known. Kallioniemi et al. (4,5) reported visual detection of 1) single deletion of a region spanning apparently 10 Mbase of 13q; 2) a more than 50-fold amplification of a 300-kb region around the *myc* oncogene on 8q; 3) *erbB2* oncogene amplification where the amplified region on 17q was estimated to be a few hundred kb and the amplification was known to be eightfold; and 4) fivefold *bcl1* amplification. Joos et al. (3) reported detection of somewhat smaller amplified regions. We have attempted to explore the detectability limit by simulation. By using simple models of DNA distribution in a metaphase chromosome, of hybridization "noise," and of image formation through the microscope and camera combination, we generated simulated chromosome images containing regions of known size and copy number change and the corresponding intensity profiles.

Chromosomes were modeled assuming that the total haploid genome is about 3.2 Gbase (10). Measurement of the total digitized areas of a number of metaphases showed that a typical pixel (0.1-μm spacing) contains on average 300 kb of DNA, and a typical point on a profile represented about 2 Mbase per chromatid, both assuming uniform distribution of the DNA across the chromosomes. DNA hybridization concentration at each pixel was assumed to be proportional to copy number but otherwise was modeled as an underlying uniform process with additive Gaussian noise. For simplicity, the normal hybridization was assumed to be noise free; therefore, the noise assumed for the test DNA hybridization corresponded directly with the noise in the measured CGH ratios at each pixel. An amplified or deleted region of given size and copy number change was constructed as a set of an appropriate number of contiguous pixels whose brightness (before addition of noise) was an appropriate multiple of the 2 copy brightness level. The effect of diffraction limited imaging was obtained by convolving the simulated image with an approximation of the point spread function of the ×63 1.3-na objective. Profile construction in the simulation was effectively identical to that described for actual data. Average profiles were obtained from multiple independent simulations. Detectability was determined subjectively, by viewing individual simulated chromosomes on a computer display, and by visually analyzing both the individual and averaged profiles.

RESULTS

Variability of Ratio Measurements and the Effect of Profile Averaging

It can be seen from Tables 1 and 2 that the normalization and profile averaging resulted in whole-arm ratios whose mean was very close to 1.0 for each chromosome. The overall mean of the ratio profiles for all chromosomes in the direct labeling experiment (Fig. 8, Table 1) was 0.984, whereas, for indirect labeling (Fig. 9, Table 2), the overall mean was 0.996. Substantial variability can be seen in Figure 9, mostly at heterochromatic regions (1 and 9 centromeres, 15 short arm), where it may be ex-

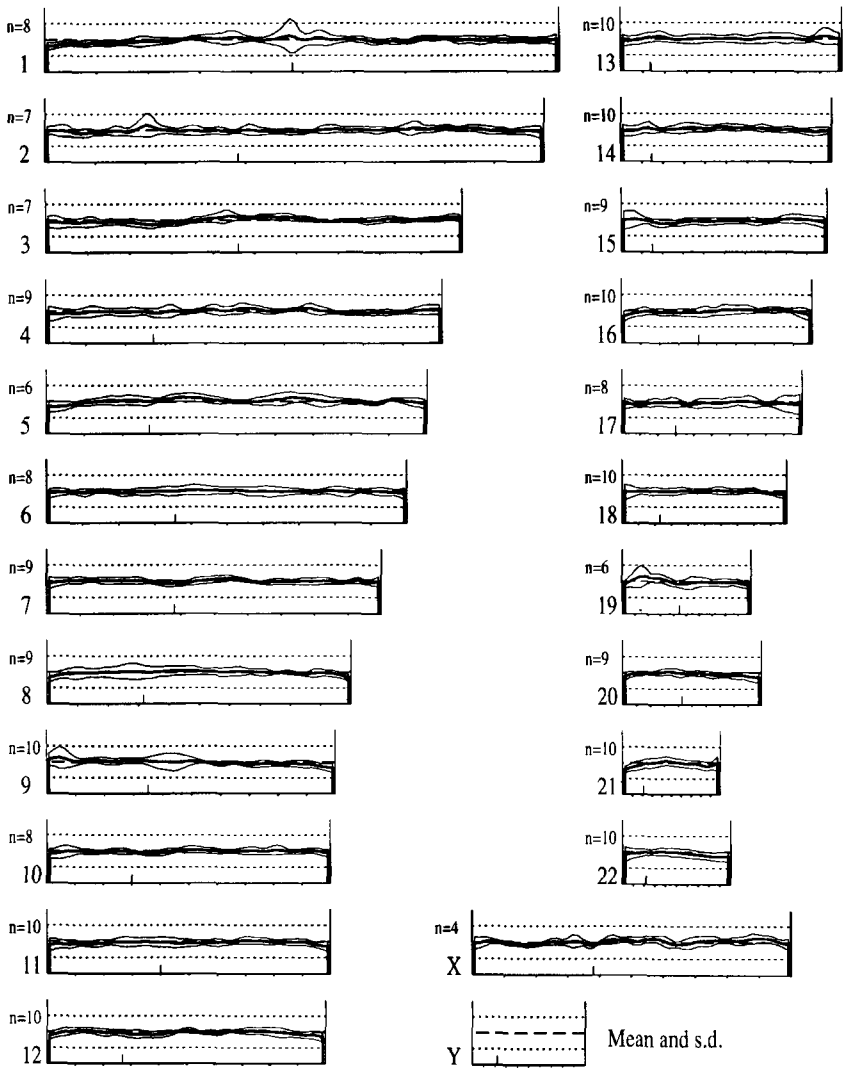

FIG. 9. Mean and standard deviation copy number karyotype profiles of normal female vs. normal female CGH analyses, pooled from five experiments using indirectly labeled DNAs obtained from a single reference sample and five separate test samples.

plained by the typically weak hybridization at these regions, which, for that reason, were excluded from the tabulated data. Other variable regions (e.g., the middle of 2p, 19p) are probably random occurrences. There are no regions in which the mean ratio is clearly and consistently in both experiments different from one. Generally, σ is less than $\underline{\sigma}$, implying that averaging has been successful in reducing random noise along the length of the profiles. In Table 1, σ' (the variability of the average ratios for individual chromosomes) is relatively small, implying only a small contribution to variability by structured noise; in Table 2, it is substantially larger for most chromosomes, possibly reflecting the fact that Table 2 was derived from five separate CGH experiments.

In practice, there are two ways to compute average ratio profiles from multiple copies of the same chromosome: 1) by length-normalizing and then averaging the individual test and reference intensity profiles, and then taking the ratio of these averages, and 2) by first calculating the ratio profile for each chromosome, then normalizing their lengths, and then taking the average of the ratio profiles. Tables 1 and 2 also show the overall mean μ^* and standard deviation σ^* of the average ratio profile $\mathbf{P}^*(i)$ for each chromosome arm calculated by method 1, which may be compared with μ and σ calculated by method 2. It can be seen that the means μ and μ^* are almost identical, whereas (contrary to our expectation) σ^* is not consistently less than σ. Overall, the root mean square difference between average profiles calculated by the two methods was 0.034 in Table 1 and 0.057 in Table 2; these are comparable in value to the corresponding σ and σ^*. We conclude that the two methods give similar results; we have chosen to use method 2, because it is a natural extension of the individual presentation of data

Table 2
Same Data as in Table 1 for CGH Experiments Comparing Five Normal Female DNAs vs. a Sixth Normal Female DNA

		Short arm						Long arm					
Chromosome	N	μ	σ	σ̲	σ′	μ*	σ*	μ	σ	σ	σ′	μ*	σ*
1	8	0.97	0.07	0.11	0.05	0.94	0.05	0.98	0.03	0.08	0.05	0.97	0.06
2	7	0.98	0.05	0.11	0.10	0.96	0.04	1.00	0.04	0.10	0.04	0.98	0.02
3	7	0.96	0.07	0.11	0.05	0.93	0.03	1.03	0.05	0.08	0.03	1.02	0.03
4	9	0.97	0.02	0.09	0.09	0.95	0.03	1.01	0.05	0.10	0.05	1.00	0.04
5	6	0.99	0.06	0.11	0.06	0.97	0.07	1.04	0.05	0.10	0.07	1.05	0.03
6	8	0.98	0.02	0.09	0.05	0.98	0.03	1.00	0.02	0.09	0.07	1.01	0.04
7	9	1.00	0.04	0.09	0.05	1.02	0.02	1.00	0.04	0.08	0.04	0.99	0.04
8	9	0.99	0.03	0.11	0.14	0.95	0.03	1.00	0.03	0.10	0.07	1.03	0.04
9	10	1.06	0.03	0.09	0.11	1.05	0.02	0.95	0.03	0.07	0.04	0.96	0.04
10	8	0.98	0.02	0.07	0.07	0.97	0.03	1.00	0.03	0.08	0.04	0.98	0.04
11	10	0.96	0.03	0.08	0.08	0.94	0.03	0.97	0.03	0.09	0.06	0.95	0.02
12	10	1.00	0.04	0.08	0.07	1.01	0.04	0.98	0.05	0.10	0.04	0.97	0.04
13	10							1.02	0.03	0.10	0.08	0.99	0.02
14	10							1.03	0.04	0.10	0.04	1.04	0.03
15	9							0.97	0.04	0.10	0.05	0.93	0.02
16	10	0.96	0.02	0.06	0.09	0.95	0.03	1.02	0.02	0.08	0.06	0.97	0.04
17	8	0.95	0.04	0.08	0.07	0.93	0.03	0.96	0.03	0.09	0.11	0.95	0.02
18	10	0.99	0.01	0.05	0.10	0.97	0.01	1.00	0.03	0.06	0.06	1.00	0.01
19	6	1.06	0.08	0.14	0.13	1.08	0.05	0.96	0.02	0.08	0.10	0.96	0.07
20	9	1.00	0.02	0.06	0.07	0.98	0.02	0.95	0.03	0.09	0.04	0.99	0.03
21	10							0.99	0.05	0.10	0.10	0.97	0.03
22	10							0.95	0.06	0.10	0.08	0.98	0.01
X	4	0.98	0.05	0.09	0.02	0.98	0.05	1.06	0.05	0.10	0.07	1.02	0.06

accumulated from many homologues (Fig. 7) and for ease of calculating the statistical measures of variability of the data described above.

Validation of Copy Number Estimation From Profiles

Figure 10a plots the measured overall mean ratios against cytogenetically determined copy number for both sets of data. To demonstrate the effect of background correction on hybridization intensity values, the analyses were repeated without background correction (Fig. 10b). Also shown are the linear regressions of ratio against copy number obtained from the set of individual chromosome mean ratios. Background correction as described resulted in slope 0.46 (desired value 0.5) and intercept 0.09 (desired value 0.0). Background correction is clearly essential in order to obtain ratio values that are directly proportional to copy number.

Simulation Study of Potential Sensitivity

Examples of the simulated images and profiles are shown in Figure 11, where the standard deviation of the additive noise was assumed to be 20% of the copy number 2 hybridization level. The simulation showed that, if the model assumptions are valid, then the limit of detectability of an amplified region in a single target chromosome could be expressed in the following terms: the product of the size of the amplified region and the excess number of copies, i.e., $L(N - 2)$, where L is the size and N is the copy number, should be greater than about 2 Mbase. This is consistent with the results reported by Kallioniemi et al. (4,5) and corresponds closely with the estimate of Joos et al. (3). The visibility threshold was to some extent dependent on the level of hybridization "noise" assumed (data not shown). Interestingly, the visibility threshold of the individual profiles was effectively the same as for the images (Fig. 11). Averaging ten profiles reduced the threshold by a factor of two; however, it must be noted that, in this simulation, unlike the case with actual experimental data, all of the profiles were accurately aligned for averaging. The simulation study also showed that, largely on account of diffraction smear, an accurate estimate of the degree of amplification could not be made from the profile height if the size of the amplified region was less than about 8 Mbase.

DISCUSSION

CGH is a valuable tool for sequence copy number analysis of whole genomes. Quantitative analysis of CGH images requires dealing with many complex issues. The method described in this paper involved adopting a number of simplifying assumptions about the nature of the primary image data, developing appropriate algorithms, and then evaluating the results from well characterized data sets to determine the degree of validity of the assumptions. The result is an analytical package that offers substantial information beyond that available by visual inspection of the hybridizations but, as will be discussed below, clearly has several avenues ripe for improvement.

The basic assumption underpinning our analysis is that the in situ hybridization reactions of the test and reference DNAs are independent, so the ratio of the binding of the two DNA samples is proportional to the ratio of the concentrations of sequences in the two samples that bind to the particular locus. The binding efficiency at the locus may be modified by local factors such as chromatin

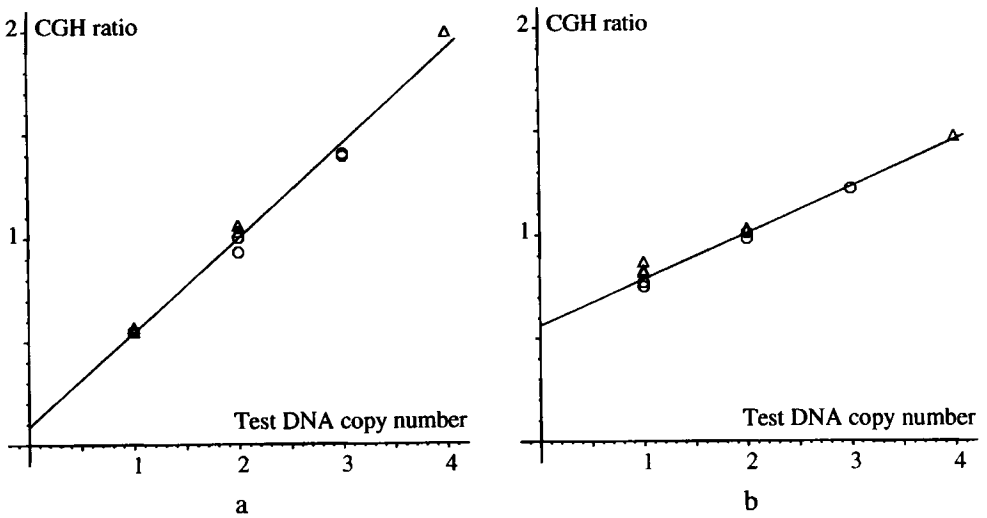

FIG. 10. Dependence of measured mean intensity ratios to known test genome copy number together with the linear regressions of the set of individual chromosome measurements. **a:** With background corrected as described in the text. **b:** Without background correction. Each mean ratio was based on between four and ten individual chromosome measurements; the standard errors of the means varied between 0.01 and 0.08 in the uncorrected data (b) and between 0.01 and 0.04 in the corrected data (a). Indirectly labeled hybridization data from the multiple X cell lines are labeled ○, whereas directly fluorochrome-conjugated data from the 600MPE cell line are labeled △.

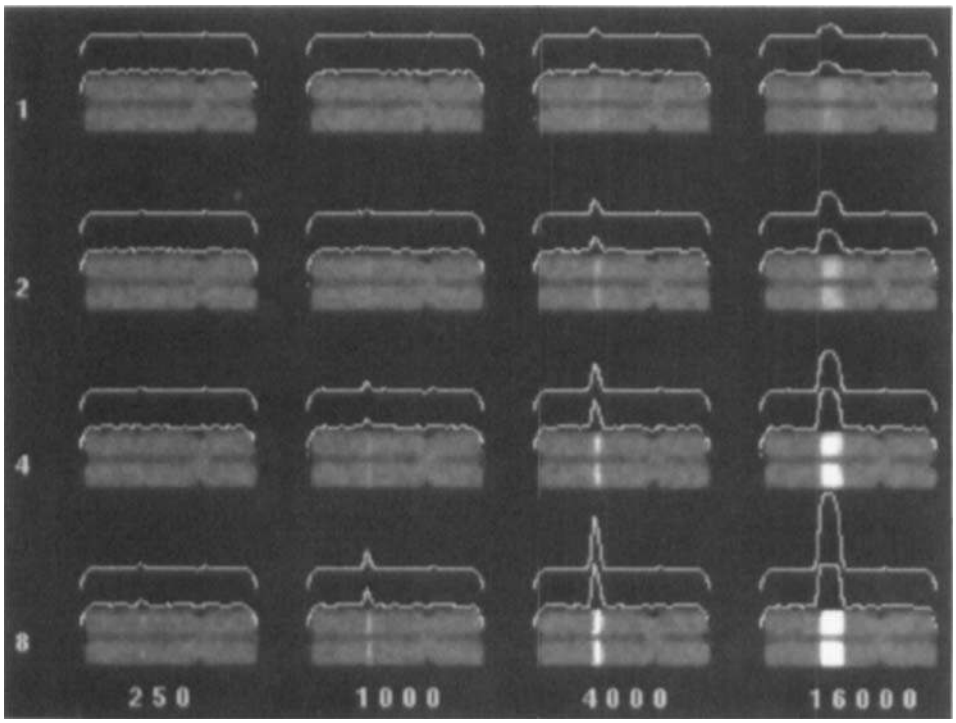

FIG. 11. Examples of simulated metaphase chromosomes showing small hybridization domains (size in kbase, horizontal axis) with the copy number increases shown (vertical axis); the corresponding ratio profiles; and (top) average profiles from ten independent simulated chromosomes.

conformation, but this will affect both DNAs equally, leaving the ratio unaltered. Furthermore, because all target chromosomes of a given type contain the same sequences at the same relative locations, essentially identical results should be obtained on each, subject only to random fluctuations. The actual value of the binding ratio is dependent both on the relative total amounts of DNA from the two genomes that are included in the reaction,

which is not itself of interest, and on the locus-specific copy number ratio of the sequences in the two genomes, which is the desired information from the experiments. Thus, the actual value of the binding ratio at a locus does not directly give the relative copy number of the sequences in the two genomes, but a change in the binding ratio from one locus to another indicates a change in the relative copy number of the two genomes between those loci.

The binding ratio of sequences in the two genomes is assessed by measuring the fluorescence intensities of the labels carried by the two genomes. These intensities are related to the binding by instrumental factors, labeling efficiency, bleaching history, etc., so that the fluorescence intensity ratio is assumed only to be proportional to the binding ratio with the proportionality factor being constant for any microscope field. Changes in the measured fluorescence ratio will then be proportional to changes in the binding ratio and, ultimately, to the copy number ratio in the two genomes. However, the proportionality constant may differ from field to field due, for example, to bleaching of the fluorochrome prior to image acquisition or to variation in camera exposure times; and it will certainly differ from experiment to experiment with the same genomes due to labeling differences, amount of DNA used, etc. Thus, we normalize the measured fluorescence ratios for each target metaphase so that the median value is approximately 1.0 (this is discussed further below). Normalized data from multiple target metaphases can be combined to reduce measurement noise.

Below, we evaluate our methodology in light of the experiments reported, and we consider other aspects of the current practice and future development of CGH analysis.

Hybridization Quality

It has been our experience that the variability of the ratio profiles depends on hybridization "quality"; if the hybridization signal was dim or had a substantially granular appearance, then the ratio profiles would appear to be noisy. Digital image analysis cannot (and should not attempt to) compensate for inadequate hybridization quality. The criteria used for assessing the quality of CGH preparations are described elsewhere (7).

Intensity Normalization and Background Correction

CGH hybridizations involve the entire genome, and the fluorescence intensity at any one position is frequently low compared to that of centromere probes or whole chromosome probes. Thus, background fluorescence, partly resulting from relatively constant instrument factors such as objective lens autofluorescence, can become a significant proportion of the signal especially when directly conjugated nucleotides are used (Fig. 2b); background correction is therefore essential. This was done in two stages. First, a global gray-level threshold in a chromosome definition image was set at a value at which local gradients were high and used to segment the entire metaphase from background. The overall relative intensities of the test and reference signals within the segmentation mask above the corresponding mean background levels were used as intensity normalizing factors. Second, for each chromosome analyzed, local backgrounds of test and reference signal were estimated and subtracted from pixel intensities. These were then normalized using the previously determined factors. The normalization procedure aimed to make background-corrected ratio values the same for corresponding loci in separate metaphases hybridized with the same test and reference DNAs, the intent being to permit valid comparison and averaging of such data. If the test and reference DNAs are identical or similar, then the average normalized ratio value should be 1.0. From Tables 1 and 2 and Figures 8 and 9, it can be seen that the background correction and normalization were highly effective in such cases.

In test DNAs obtained from tumors, there is usually a well-defined modal genomic copy number possessed by the majority of chromosome loci, in which case the computed ratio profile value at such loci was also near to 1.0 (as shown, e.g., in Figs. 6, 7). Our use of the median rather than the mean pixel value as the basis of normalization was aimed at producing this result in cases where the aneuploid regions of the genome were unevenly distributed around the modal copy number; in such cases, the mean will bias the average ratio away from 1.0, the median less so. However, it is quite possible to have a genome in which half of the chromosomes are present at four copies, whereas the remainder have two copies, in which case a ratio of 1.0 would represent none of the genomic regions present in the test DNA. In solid tumors, the genome may be heterogeneous, and the sample may contain a proportion of normal cells. In such cases, a ratio of 1.0 corresponds to the *average* copy number of the complex mixture of genomes in the specimen.

Background correction, as described, has assumed that the interchromosome background is also additive "under" each chromosome. This assumption needs to be checked in the future, but there definitely is an additional component of background that has not been explicitly corrected, that due to incomplete suppression of hybridization from repeat sequences. Chromosome painting by whole chromosome probes provides an appropriate analogy and shows a contrast ratio in the region of 5:1–10:1 between signal and unsuppressed background (11). It is therefore likely that between 10% and 20% of the test and reference signals at any locus may be due to such repeats even after background subtraction as described.

The repetitive sequences causing this background are distributed throughout the genome (although somewhat nonuniformly, because hybridizations with them produce banding patterns in chromosomes). Thus, to a first approximation, the total amount of repeat sequence DNA in a test or reference sample is a constant fraction of the total DNA independent of copy number variations that might be present. Therefore, after intensity normalization, we expect that the contributions of unsuppressed

repeats to the test and reference signals will be equal at any particular location x along the chromosome profile. This common contribution, ε_x, is, however, expected to vary with position.

The signal due to the unique sequences may also vary with position due to target related factors in addition to any possible copy number change. Thus, after intensity normalization, the test signal at position x due to unique sequences is $\xi_x t_x/\mathbf{t}$, where t_x is the test DNA sequence copy number at x, $\mathbf{t}$ is the mean copy number in the test sample, and ξ_x is local hybridizability at x. The corresponding normalized reference signal is just ξ_x. (Note: this assumes that the reference genome has constant copy number. What follows may need to be modified for the Y chromosome if male DNA is used as a reference or, more substantially, if a complex reference genome is used.) Therefore, the measured intensity ratio $\rho_x = (\xi_x t_x/\mathbf{t} + \varepsilon_x)/(\xi_x + \varepsilon_x)$. We note several consequences of this relationship:

1) If the test genome has constant copy number $t_x = \mathbf{t}$ throughout, e.g., a normal genome as was used in some of the tests in this paper, or a tetraploid genome, then ρ_x is a constant independent of the degree of suppression of the repetitive sequences.

2) If $\varepsilon_x << \xi_x$, that is, the repetitive sequences are substantially suppressed, then to first order

$$\rho_x = t_x(1 - \varepsilon_x/\xi_x)/\mathbf{t} + \varepsilon_x/\xi_x. \quad (3)$$

In these circumstances, ρ_x is linear in copy number t_x but has slope $<1/\mathbf{t}$ and a nonzero intercept. For example, the linear regression of the data in Figure 10a had an intercept of 0.09, implying that the average of ε_x/ξ_x over the chromosomal regions measured in the experiments was 0.09. Taking this value of ε_x/ξ_x and the prior knowledge that, in this experiment, $\mathbf{t} \approx 2$, the prediction for the slope is $(1 - 0.09)/2 = 0.455$, which agrees very well with the measured value of 0.46 and provides some confirmation for our model. Because of intensity normalization, the average values $\bar{\xi}$ of ξ_x and $\bar{\varepsilon}$ of ε_x are related by $\bar{\xi} + \bar{\varepsilon} = 1$. Thus, we can estimate from the data that $\bar{\varepsilon} = 0.08$, indicating that the repetitive sequences in these hybridizations contribute less than 10% of the total signal.

3) Again assuming that ε_x is small, equation 3 shows that, in regions of the test genome where the copy number t_x differs from the average $\mathbf{t}$, the error in the measured ratio is $\varepsilon_x(1 - t_x/\mathbf{t})/\xi_x$. For example, in a tetraploid region of a basically diploid cell, the error is $-\varepsilon_x/\xi_x$. We predict that, where locally constant t_x is substantially different from $\mathbf{t}$, fluctuations in the ratio error as ε_x/ξ_x changes with position may become visible as a consistent banding pattern in the ratio profile.

Quantitative Accuracy

We have tested the combined hybridization and quantitative analysis system to see to what extent the analytical approximations and the reality of the actual hybridization and antibody staining regimes permit quantitative interpretation of ratio changes. Other factors that can reduce the accuracy of ratio estimation include "crosstalk" between the fluorochrome channels and chromatic aberration of the incident illumination. In preliminary experiments, the relative intensity of a Texas red labeled probe when measured in the FITC channel was less than 2% (2); FITC leakage into the Texas red channel was too low to be measured. It would be possible to make a correction using the method described by Castleman (1). However, in view of the smallness of any bias resulting from uncorrected crosstalk with the currently used labels, we chose not to do so. Similarly, the ratio between the excitation intensities for Texas red and FITC was measured to vary by about ±2% across the diameter of a metaphase (2); again, this effect, which contributes a small random variation of the ratios measured on homologous chromosomes, was regarded as being too small to warrant correction. The results presented in Figures 8–10 and Tables 1 and 2 demonstrate that, in practice, these omissions were justified. In particular, Figure 10a shows that, overall, the proportionality between copy number and measured ratio is good even between experiments, so long as the modal test DNA copy number is two. If further calibration experiments confirm that the level of unblocked, nonspecific hybridization is relatively stable, it will be possible in the future to factor it out of measured ratios.

Although the problem did not arise with the any of data presented here, we have, on occasion, noted regions of the genome that gave anomalous ratios when indirect fluorochrome labeling was used, in particular, across much of chromosome 19 and distal 1p. We note, first, that these are particularly pale staining regions in G-banding; second, that the effect may be controlled for by a reversed labeling experiment in which the labels applied to test and reference DNA (e.g., biotin and digoxigenin) are reversed. This should result in identical ratios of test to reference DNA at each locus as did the first hybridization; if it does not, then a hybridization artefact should be suspected.

Finally, physical limitations such as diffraction prevent quantitative interpretation of CGH ratio changes in very small regions of amplification or deletion. Simulation indicated a lower limit for ratio quantification of around 10 Mbase, but factors such as disruption of chromosome structure by denaturing and hybridization may make this a considerable underestimate.

Profile Averaging

Because the hybridization data on all copies (and on each chromatid) of a given chromosome type should be the same following normalization and background correction, averaging the data over chromatids and multiple homologues has the potential to reduce the noise and thereby increase the accuracy of the analysis. Correct averaging requires alignment of corresponding loci on the chromosomes. Our procedure was based on identifying the ends of the chromosomes only and assumed linear stretching for normalizing chromosome lengths. Errors in establishing the chromosome ends in the digitized image, centromeric polymorphisms, and nonlinear

contraction of metaphase chromosomes all contributed to imperfect alignment. Alignment problems are apparent in Figure 7a, chromosome 11, which has a deletion on the long arm. Work is underway to develop a procedure to align chromosomes on the basis of matching the counterstain bands.

Ratio profile averaging can be performed either by averaging test and reference intensity profiles individually and then taking the overall ratio or by averaging the individual ratio profiles. These should give equivalent results except possibly at loci where the reference signal is weak resulting in lower precision of the ratio estimate. Tables 1 and 2 confirm that there is little difference in practice. The second method has the great advantage that a simple method for computing confidence limits is available, and is, therefore, our method of choice.

Averaging reduces both random noise and structured noise due to positional effects of chromosomes within metaphases. A reduction of random noise can be seen in the generally lower values of σ compared with $\underline{\sigma}$ in Tables 1 and 2. The variation due to structured noise is a major contributor to the values σ' in Tables 1 and 2; the effect on the mean profile should generally reduce by a factor $\sqrt{n}$, where n is the number of profiles included in the average.

Objective Criteria for Interpreting Nonmodal Copy Number

Objective criteria are required for judging whether a particular region has variant copy number; we propose the following. CGH analysis, as we have implemented it, establishes a mean test to reference profile intensity ratio of 1.0 for the mean copy number of the test genome. Deviations from this mean copy number may be scored by inspection of the 95% or 99% confidence limits of averaged ratios plotted in the composite karyotype; this has been automated. Figure 7b shows thin underscoring for regions that are significantly below the mean copy number and thick underscoring for regions of significantly increased copy number. A similar approach may be taken to the more general question of whether two parts of the genome have different copy number; here, conventional statistical techniques for testing the nonidentical distribution of two sets of samples may be applied on a point-wise basis. Such a method ignores the additional factor that nearby points on a ratio profile are clearly dependent in the statistical sense, and we intend to investigate statistical techniques that incorporate this contextual information.

Limiting Sensitivity of CGH Analysis

It is not known just how small an amplified region can be detected by the CGH imaging techniques. Significant factors are likely to include the degree of amplification, the size of the affected region, and the contraction of the target metaphases. Using a plausible model that takes account of the likely variability or "noise" in the hybridization process as well as the physical characteristics of the imaging system, simulation experiments have shown that a copy number increase of 50% (i.e., one additional copy in a basically diploid DNA sample) should be visible if the amplified region is about 2 Mbase or larger; for a 0.25 Mbase amplification to be visible, the copy number increase needs to be about 400% (Fig. 11). These simulation results are in close agreement with Joos et al. (3), who estimated that CGH can detect an amplified region as small as 100 kbase as long as the copy number has increased by at least 20 (i.e., a 1,000% amplification). All of these results point to a lower limit of detectability that is determined by the product of the excess copy number and the size of the amplified region; the detectability threshold of this product appears to be in the vicinity of 2 Mbase if metaphase chromosomes are used as the target. In the case of deletions in near-diploid DNA, the maximum copy number loss is, of course, two (or 100%), and the simulation showed that total deletions smaller than 1 Mbase or 50% (single copy) deletions smaller than about 2 Mbase are unlikely to be resolved. Practical experience has yet to confirm these predictions.

Simulation results should also be regarded as preliminary, because verification of some of the particular numerical assumptions in the model was not possible; and, in terms of DNA condensation, they apply only to metaphase chromosomes. Different results might reasonably be expected if prometaphase preparations were used as the target. The real appearance of small regions of amplification or deletion will not usually be as simple as in the simulation model. Furthermore, aneuploidy, intratumor heterogeneity and presence of normal cells in tumor specimens will significantly affect sensitivity. Finally, as noted above, the measurement of ratio cannot be expected to be linear with copy number for small regions comparable in size with the optical resolution limit, and this was borne out by the simulation experiment.

Thus, overall, simulation provides a useful guide to sensitivity but awaits experimental confirmation. This could, for example, be based on a set of cell lines carrying small regions of amplification or deletion whose location and size has been well characterized by alternative techniques (V. van Heyningen, personal communication).

Future Directions for Improved CGH Analysis

The simulation and the reduction of variability in the normal test DNA vs. normal reference experiments (Tables 1, 2) both showed the benefits of averaging a number of chromosome profiles. However, the difficulty of precisely locating the chromosome endpoints, "slippage" between chromatids, differential contraction effects along the chromosomes, centromeric polymorphisms, and measurement artefacts, such as not-quite-ideal positioning of the medial axis, can all contribute to a failure of the profiles to align perfectly among a set of homologous chromosomes (see, e.g., chromosome 11 in Fig. 7a). Under such conditions, profile averaging merely smears the signal. What is required instead is independent alignment of the chromosome landmarks (telomeres, centromere, and major bands) computed on the basis of the counterstain image and its profile, which can then be applied to

the ratio profiles. For the smallest regions of amplification or deletion, profiling itself may smear the signal, and averaging the two-dimensional images would most likely have similar problems. Instead, we propose that *recognition* of common signals present in a substantial proportion of multiple copies of a chromosome should be used, much as one might recognize small unique copy probes hybridized to metaphases. Such a technique would be particularly appropriate to high factor amplifications of small extent.

CGH analysis is labor intensive in terms of the time spent interacting with the system in order to carry out the tasks of segmentation, chromosome identification, rejection of badly measured chromosomes, etc. Profile averaging multiplies the cost by the number of cells deemed necessary. In order to reduce these costs, more fully automatic analysis incorporating automatic digitization, segmentation, classification, and profile measurement is required followed, finally, by operator review and selection of adequately preprocessed data.

We expect that the demand for CGH will continue to grow rapidly. It is likely that the preparations will improve as FISH techniques and reagents are further developed, and we intend to pursue the image analysis goals actively in order to take full advantage of this new technology.

ACKNOWLEDGMENTS

We thank Sandy Devries, Gayatry Mohapatra, and Rick Segraves for specimen preparation and assistance with analyzing the data, and Jim Mullikin for preparing Figure 3. Additional support was provided by the EC Concerted Action on Automation of Molecular Cytogenetic Analysis (Project no. PL921307).

LITERATURE CITED

1. Castleman KR: Color compensation for digitized FISH images. Bioimaging 1:159–165, 1993.
2. Hill W, Sudar D, Perry P, Piper J: Digital image analysis techniques for measuring the performance of fluorescence microscopes. Genetics Unit, Research Report RN94-005, MRC Human Genetics Unit, Edinburgh, 1994.
3. Joos S, Scherthan H, Speicher MR, Schlegel J, Cremer T, Lichter P: Detection of amplified DNA sequences by reverse chromosome painting using tumor DNA as probe. Hum Genet 90:584–589, 1993.
4. Kallioniemi A, Kallioniemi O-P, Sudar D, Rutovitz D, Gray JW, Waldman FM, Pinkel D: Comparative genomic hybridization for molecular analysis of solid tumors. Science 258:818–821, 1992.
5. Kallioniemi O-P, Kallioniemi A, Sudar D, Rutovitz D, Gray JW, Waldman FM, Pinkel D: Comparative genomic hybridization: A rapid new method for detecting and mapping DNA amplification in tumors. Semin Cancer Biol 4:41–46, 1993.
6. Kallioniemi A, Kallioniemi O-P, Piper J, Tanner M, Stokke T, Chen L, Smith HS, Pinkel D, Gray JW, Waldman FM: Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. Proc Natl Acad Sci USA 91:2156–2160, 1994.
7. Kallioniemi A, Kallioniemi O-P, Piper J, Isola J, Waldman FM, Gray JW, Pinkel D: Comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. Genes Chromosom Cancer 10:231–234, 1994.
8. Kallioniemi O-P, Kallioniemi A, Mascio L, Sudar D, Pinkel D, Deaven L, Gray JW: Physical mapping of chromosome 17 cosmids by fluorescence in situ hybridization and digital image analysis. Genomics 20:125–128, 1994.
9. du Manoir S, Speicher MR, Joos S, Schrock E, Popp S, Dochner H, Kovacs G, Robert-Nicoud M, Lichter P, Cremer T: Detection of complete and partial chromosome gains and losses by comparative genomic in situ hybridization. Hum Genet 90:590–610, 1993.
10. Morton NE: Parameters of the human genome. Proc Natl Acad Sci USA 88:7474–7476, 1991.
11. Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray J: Fluorescence in situ hybridization with human chromosome-specific libraries: Detection of trisomy 21 and translocations of chromosome 4. Proc Natl Acad Sci USA 85:9138–9142, 1988.
12. Piper J, Granum E: On fully automatic feature measurement for banded chromosome classification. Cytometry 10:242–255, 1989.
13. Piper J, Rutovitz D: Data structures for image processing in a C language and Unix environment. Pattern Recogn Lett 3:119–129, 1989.
14. Rutovitz D: Expanding picture components to natural density boundaries by propagation methods. The notions of fall set and fall distance. Kyoto, Japan: Proc IVth IJCPR, 1978, pp 657–664.
15. "SCILimage Manual." Amsterdam: University of Amsterdam, 1992.