Genomic microarrays in human genetic disease and cancer

Donna G. Albertson[1,2,*] and Daniel Pinkel[2]

[1]Cancer Research Institute and [2]Department of Laboratory Medicine, Box 0808, University of California San Francisco, San Francisco, CA 94143-0808, USA

Alterations in the genome that lead to changes in DNA sequence copy number are a characteristic of solid tumors and are found in association with developmental abnormalities and/or mental retardation. Comparative genomic hybridization (CGH) can be used to detect and map these changes. Recent improvements in the resolution and sensitivity of CGH have been possible through implementation of microarray-based CGH (array CGH). Here we discuss the performance characteristics of different array platforms and review some of the recent applications of array CGH in cancer and medical genetics.

INTRODUCTION

Comparative genomic hybridization (CGH), a technique that detects and maps changes in copy number of DNA sequences, has been widely used for the analysis of tumor genomes and constitutional chromosomal aberrations since it was first reported by Kallioniemi and colleagues in 1992 (1). In CGH, DNA from a test (e.g. tumor) and a reference genome (genomic DNA from a normal individual) are differentially labeled and hybridized to a representation of the genome, which was originally a metaphase chromosome spread. Hybridization of repetitive sequences is blocked by the addition of Cot-1 DNA. The fluorescence ratio of the test and reference hybridization signals is determined at different positions along the genome and provides information on the relative copy number of sequences in the test genome compared with a normal diploid genome. In the past few years, microarray-based formats for CGH (array CGH) have been reported and are beginning to be widely used in preference to chromosome-based CGH. As discussed below, arrays made from large genomic clones and cDNAs have been used most often for this purpose. The array format for CGH can provide a number of advantages over the use of chromosomes, including higher resolution and dynamic range, direct mapping of aberrations to the genome sequence and higher throughput. Furthermore, since the array format lends itself to automation, array CGH-based *in vitro* diagnostic devices are possible. Here, we discuss the platforms and performance requirements for different uses of array CGH and review recent applications in cancer and human genetic disease.

ARRAY CGH PLATFORMS AND PERFORMANCE REQUIREMENTS

A variety of chromosomal aberrations underlie developmental abnormalities and cancer (Fig. 1). Aberrations leading to gains or losses of part of the genome can be detected by CGH and include interstitial deletions and duplications, non-reciprocal translocations and gene amplifications. Aberrations that do not result in copy number changes can often be detected by other techniques such as chromosome banding, SKY or M-FISH and loss of heterozygosity or allelic imbalance (LOH). It is important to note that CGH does not provide information on ploidy or location of the rearranged sequences responsible for the copy number change. Furthermore, the capability of array CGH to detect aberrations spanning small regions of the genome depends on both the size and spacing of the clones on the array.

Different applications of array CGH impose different performance requirements, so that certain approaches may be suitable for particular applications, while others are not. The first consideration is complexity of the genome that is being analyzed. As one moves from organisms with small genomes, such as yeast, to mammals with large genomes, the measurements become more difficult because of the decreasing partial concentrations of each portion of the sequence that is involved in the hybridization to the array elements. The easiest task is detection of large increases in copy number in DNA extracted from homogeneous cell lines. Achieving adequate performance is more difficult if one desires to reliably detect low level (single copy) gains and losses, especially as the size of the aberrant region decreases. Another dimension of challenge

*To whom correspondence should be addressed. Tel: +1 4155028463; Fax: +1 4154768218; Email: albertson@cc.ucsf.edu
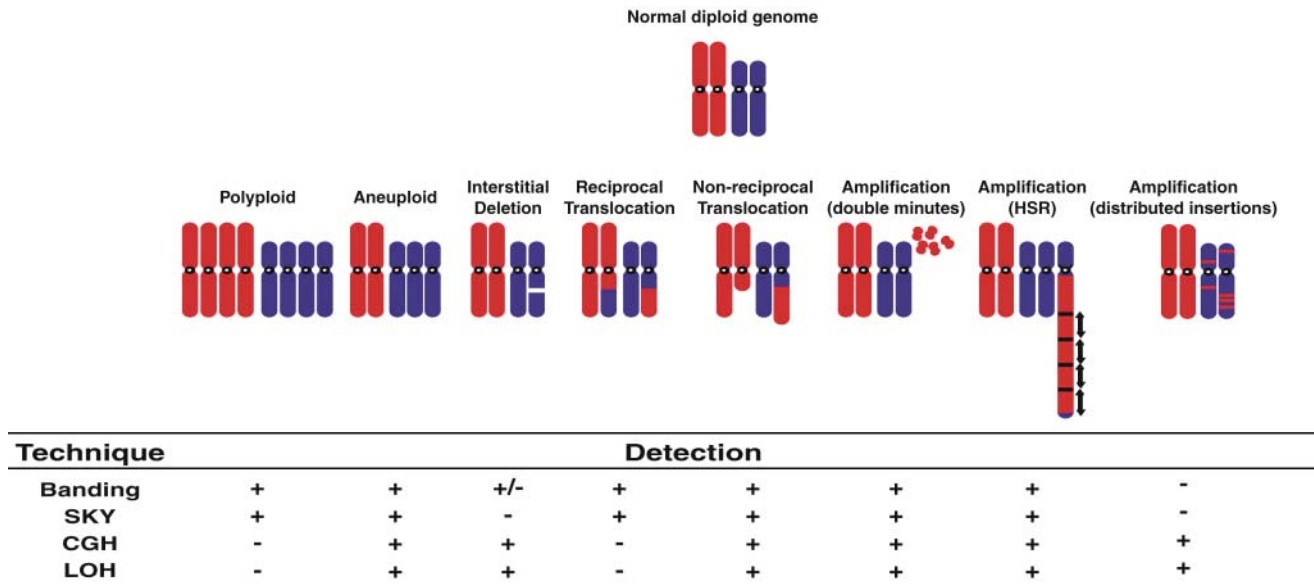
Normal diploid genome

Technique	Polyploid	Aneuploid	Interstitial Deletion	Reciprocal Translocation	Non-reciprocal Translocation	Amplification (double minutes)	Amplification (HSR)	Amplification (distributed insertions)
Banding	+	+	+/-	+	+	+	+	-
SKY	+	+	-	+	+	+	+	-
CGH	-	+	+	-	+	+	+	+
LOH	-	+	+	-	+	+	+	+

Figure 1. Detection and mapping of chromosomal aberrations by different molecular and cytological techniques. Rearrangements involving two chromosomes of a diploid genome are shown. The ability of various techniques to detect the aberration is indicated in the table. Modified from Albertson *et al.* (67).

involves the use of tissue specimens, which may contain heterogeneous cell populations, for example genomically normal cells within tumors. More difficult still is the use of formalin-fixed archival tissue. Moreover, the use of tissue from clinical specimens may impose constraints on the amount of DNA available for analysis. Finally, different applications have different tolerances for error, which substantially affects the performance requirements. For example, if one seeks composite information on the general characteristics of aberrations that occur in a set of specimens, the penalty for any single error is small. Indeed, missing a whole type of aberration is acceptable if other valuable information is obtained. However, it is much more of a challenge to obtain specific information from an individual specimen with sufficient confidence for clinical use.

A number of different array platforms have been used for CGH measurements in mammalian genomes. The various approaches have employed large insert genomic clones, such as bacterial artificial chromosomes (BACs), cDNA clones and oligonucleotides for array spots. An analysis typically requires several hundred nanograms of genomic DNA from the specimen when using some BAC arrays (2), or one or more micrograms for cDNA arrays and some BAC arrays (3,4). Whole-genome amplification procedures have shown promise to substantially reduce the amount of specimen required (5).

The use of large insert genomic clones such as BACs for array CGH provides sufficiently intense signals so that accurate measurements can be obtained over a dynamic range in copy number from deletions to amplifications on the order of a factor of at least 1000 (6). Thus, the measurement precision is such that single copy changes affecting individual clones on the array can be detected (2) and aberration boundaries can be located to within a fraction of a length of a BAC (7). However, propagating and printing BACs can be problematic. They are single copy vectors and so yields of BAC DNA are low and

solutions of the high molecular weight DNA can be viscous, making it difficult to print. Arrays described in initial reports of array CGH used whole BACs isolated from large bacterial cultures, and the DNA was often sonicated to reduce the molecular weight (6,8,9). Since growing and processing large bacterial cultures is not viable for large arrays, a number of methods have been devised that use representations of BACs prepared by ligation-mediated PCR (2,10), degenerate oligonucleotide primed PCR (DOP–PCR) using the 6MW primer (11–13) or a modified DOP–PCR protocol (14). The goals of these methods are to provide the most complete amplification of all the sequences in the cloned DNA. If the complexity and sequence balance of the source clones is maintained, then the BAC representations are expected to yield sufficiently intense signals to perform as well as whole BACs. Indeed, the ratios obtained on arrays comprising ligation-mediated PCR BAC representations were shown to be essentially identical to ratios previously reported on DNA from the same BACs (2), whereas arrays made using solutions prepared by DOP–PCR with the standard 6MW primer were found to perform less well than whole BACs (11). However, the modified DOP–PCR protocol, which uses three different primers selected to optimally amplify human genome sequences, but not *E. coli* DNA, appears to provide improved sensitivity and reproducibility compared with amplification with 6MW. The improved performance has been attributed to both increased representation of the genome sequence and decreased contamination with *E. coli* sequences (14).

Other approaches for the preparation of genomic clone arrays were used to prepare a full-coverage array for chromosome 22. A sequence-based approach was used to design primers that amplify only non-repetitive DNA from a region of the genome. The pools of non-redundant DNA fragments were then pooled for spotting (15). Rolling circle amplification using phi29 DNA polymerase was also used to produce copies of cloned genomic

DNA, resulting in exponential amplification that yielded 50 µg of DNA from 10 ng of template DNA in a single overnight reaction. These approaches have advantages for array fabrication, including removal of repeat sequences from the genomic regions spotted on the array and potentially improved representation of cloned DNA by rolling circle amplification.

DNA copy number measurements can also be made using arrays containing spots made from cDNAs (3) or oligonucleotides. Typically these arrays were initially produced to measure gene expression. The advantages of these arrays are that they frequently contain a large number of elements because they were produced to comprehensively assess the transcriptome, DNA copy number information is obtained on the same sequences that are used to assess expression, and they may be available commercially prior to general availability of BAC arrays. The cDNA arrays have proven their ability to detect amplified sequences using several micrograms of genomic DNA isolated from cell lines and frozen tissue (3,16–19). The large copy number changes in these aberrations provided sufficient signal so that boundaries of the amplicons could be determined with high resolution because the amplification status of individual array elements could be determined. The detection of lower level copy number changes on arrays of this type requires calculating the running average of multiple clones, typically five to 10, along the genome, and frequently entails discarding measurements on a substantial fraction of clones because they do not provide adequate signals (3,17,18). Thus the actual genomic resolution of the boundaries of single copy changes and the ability to detect focal single copy changes is considerably less than implied by the average genomic spacing between the clones on the array. Detection of homozygous deletions has been reported in cell lines after statistical analysis of the data (19) (Baldocchi, R., manuscript in preparation). Utilizing the same arrays for expression analysis and DNA measurements (17,18) has permitted assessment of the relationship of mRNA expression levels to DNA copy number broadly across the genome. Similar data can be obtained from BAC arrays if one correlates the changes in expression of genes in the vicinity of a BAC with its copy number variation.

The use of oligonucleotide arrays for DNA copy number analysis is being approached in two ways. In the first, the complexities of the genomic DNA in the test and reference samples are simplified using PCR strategies (Baldocchi, R., manuscript in preparation) (20) coupled with array elements chosen to contain sequences within the portion of the genome that is amplified. This approach has the potential to produce bright signals because of the complexity reduction of the amplification processes, but the accuracy of the measurement depends critically on the similarity of the amplification efficiency of the corresponding portions of the test and reference genomes.

More recently, several groups have begun to report data at meetings on the use of oligonucleotide expression arrays for mammalian DNA measurements without complexity reduction. These results clearly show the ability to detect high-level amplifications and to determine the boundaries of high copy number portions of the genome. However, detection of single copy changes is more problematic, requiring averaging over large numbers of neighboring array elements, with a corresponding reduction in genomic resolution, in order to detect single copy changes. In the future, techniques for making and hybridizing to arrays of short sequences may improve sufficiently so that one can obtain both the measurement precision and reliability now available from certain of the BAC array techniques, as well as the genomic resolution promised by the large number or array elements on oligonucleotide arrays.

ANALYSIS OF TUMOR GENOMES

Analyses of tumor genomes by array CGH have employed arrays focused on a particular region of the genome (4,21,22), selected regions known to be frequently aberrant in tumors (23–27) or genome-wide arrays (18,28–32). In all cases, the enhanced resolution possible with array CGH compared with chromosome CGH has been demonstrated by the fact that array CGH found copy number aberrations that were not seen using chromosome CGH.

Application of genome-wide array CGH in different tumors is revealing that they differ not only in the regions that are aberrant, but also the types of copy number aberrations that are present (Fig. 2). Thus, it appears that tumor specific types of copy number profiles, or copy number phenotypes, arise due both to selection for particular changes affecting gene expression as well as different kinds of underlying genetic instability (33). For example, cytogenetic analyses have shown previously that colon tumors with defects in mismatch repair (MMR) have fewer copy number alterations than MMR-proficient colon tumors. These tumors also differ in their histology, in the genes that are inactivated and in their response to therapy. Thus, they can be readily distinguished based on the types of genetic instability displayed and on the selection of the genes that are altered. Analysis of the numbers and types of aberrations in the array CGH copy number profiles from MMR-proficient and -deficient cell lines confirmed the cytogenetic observations, but also found that MMR-deficient cell lines with alterations in *MSH2* had fewer aberrations than those with alterations involving *MLH1*. Further studies of a model system demonstrated the importance of the genetic background of cells in shaping the copy number profiles of tumor cell genomes (33).

Tumor classification based on copy number profiles obtained using array CGH has been reported for several small studies (30,31,34). In a study of liposarcoma, copy number profiles had greater power to discriminate between dedifferentiated and pleomorphic subtypes than expression profiling. A study of bladder tumors failed to find significant relationships between copy number changes and tumor stage and grade (35). However, an analysis of copy number changes at particular loci revealed that certain aberrations occurred together (e.g. gains of *ERBB2* and *CCNE1*), whereas alterations in the copy number of loci harboring genes that function in the same pathway such as gains of *CCND1* and *E2F3* were found to be 'complementary' as they did not occur in the same tumors (35). This study suggests that copy number profiles may have utility for understanding deregulation of cellular control pathways in solid tumors. It is likely that the generality of these observations will become apparent in the next year as more array CGH studies of other tumor types are published.
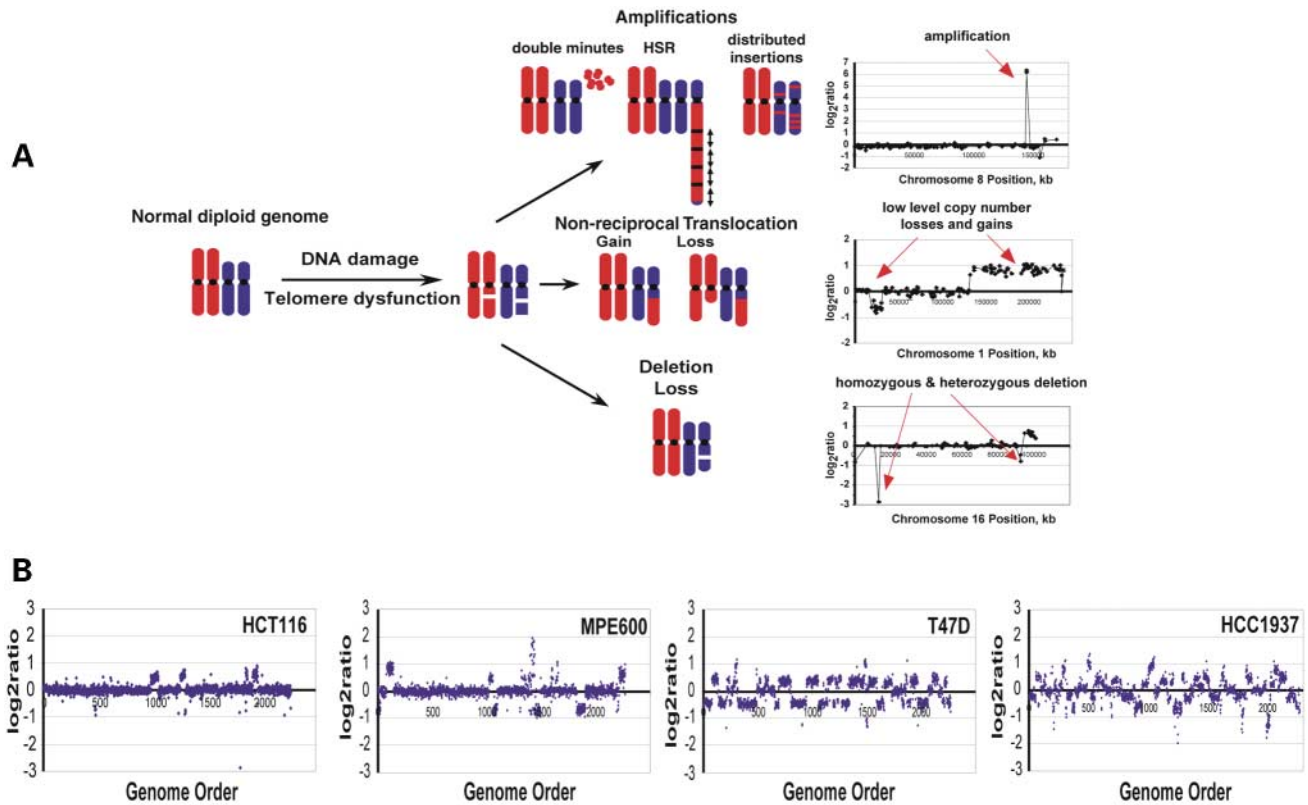
Figure 2. Detection of copy number aberrations in tumor genomes by array CGH. (**A**) Chromosomal aberrations in cancer are likely to arise following inappropriate management of DNA damage or telomere dysfunction. Common aberrations include gene amplifications, non-reciprocal translocations and interstitial deletions. Amplifications may be visible cytogenetically as double minutes, chromosomes with homogeneously staining regions (hsr) or the amplified DNA may be distributed at multiple sites. The array CGH copy number profile of the amplified *MYC* in COLO320 is shown. The amplification level is ∼70 fold ($\log_2$ratio > 6). Breakage of a chromosome or a non-reciprocal translocation event may lead to low level copy number changes, as shown in the copy number profile of chromosome 1 from 600MPE. Homozygous deletions are indicated by $\log_2$ratio < −2 and heterozygous deletions by $\log_2$ratio = ∼−1, as shown in the copy number profile for chromosome 16 in HCT116. The $\log_2$ratios are plotted on individual chromosomes according to the positions of the clones on the UCSC August 2001 freeze of the genome sequence (http://genome.ucsc.edu). Complete array data sets are available from Snijders *et al.* (2). (**B**) Whole genome array CGH copy number profiles. The spectrum of copy number aberrations seen in different tumor types varies as illustrated by the copy number profile of HCT116 (mismatch repair defective colon tumor cell line, with few copy number alterations), 600MPE (breast tumor cell line with amplification of *CCND1*), T47D (breast tumor cell line with many low level copy number changes) and HCC1937 (*BRCA1* deficient breast cancer cell line, with many copy number changes). The $\log_2$ratios for each chromosome in order from 1p to Xqter are plotted in order according to the positions of the clones on the UCSC August 2001 freeze of the genome sequence (http://genome.ucsc.edu). Complete array data sets for HCT116, 600MPE and T47D are available from Snijders *et al.* (2).

A number of studies have taken advantage of the higher resolution afforded by array CGH to more precisely map the boundaries and amplification maxima of amplified regions. Once these are known, candidate oncogenes mapping within the region or at the amplification maximum can be readily identified from the genome sequence databases. Investigation of expression levels of these candidate genes in tumors and cell lines can then be used to determine which of the candidates are most likely to contribute to the disease phenotype and to be the 'driver gene(s)' for amplification (4,7,17,18, 21,22,28,29). For example, this approach identified *CCNL1* as an overexpressed gene in well-differentiated head and neck tumors (22). It appears that this approach to candidate cancer gene identification will be fruitful, since a good correspondence was found between copy number alterations and changes in gene expression in breast cancer as measured by genome-wide array CGH and global expression profiling (17,18).

COPY NUMBER VARIATIONS IN HUMAN GENETIC DISEASE

Chromosomal aberrations are associated with a number of congenital anomalies characterized by various dysmorphologies and/or mental retardation. Currently, cytogenetic analysis of Giemsa stained metaphase chromosomes (G-banding) is applied to ascertain such abnormalities. Typically these assays resolve 550 bands resulting in the capability to identify deletions or duplications and inversions if they are of sufficient size. Application of fluorescent *in situ* hybridization (FISH) can improve resolution and is the gold standard for detection of changes in DNA copy number as well as inversions. However, it is typically applied in a targeted manner that assesses one or several candidate loci at a time, and it has been estimated that a cytogeneticist may require 5–7 h to screen each case (36). Genomic arrays offer several advantages for screening of copy number changes in human genetic

disease, including higher resolution mapping, directly to genome sequence and higher throughput due to the massive parallelism of the assay. The capability of array CGH to reliably detect copy number alterations affecting single BACs on a large array is illustrated in Figure 3A and B, which shows an example of the FISH and array CGH analysis, respectively, of a patient who tested positive for a deletion in the DiGeorge critical region on chromosome 22q11.2 by both FISH and array CGH. The deleted region is clearly evident with a log2ratio close to the expected value of −1 for a heterozygous deletion (Rauen *et al.*, in preparation). Several reports of the analysis of constitutional chromosomal aberrations using genomic arrays have appeared recently (37–39). The array data were found to be concordant with previous G-banding or FISH analysis, but they provided higher resolution definition of the involved aberrations. In some cases, array CGH revealed additional aberrations, highlighting the value of the whole genome scan as compared to FISH, which can only provide information on those loci whose status have been queried.

Small chromosomal rearrangements involving the telomeres have been found in association with idiopathic mental retardation. The frequency of such aberrations across 22 studies involving 2500 patients has been reported to be 5%, increasing to 7% in patients with moderate or severe retardation (40), although other studies have found subtelomeric rearrangements to occur more rarely (41,42). Arrays designed to detect copy number alterations in subtelomeric regions using the second generation set of human subtelomere-specific probes (42) have been reported (11,43). In a blinded study of 20 patients with known cytogenetic abnormalities, array CGH found the expected aberrations as well as additional cytogenetically cryptic copy number changes (11). Although no false negatives were reported in this study, it is important to consider the false positive rate when considering implementing array-based assays in the clinic. The thresholds for gain or loss from the expected ratio of 1 were set at 0.8 and 1.2, respectively, based on multiple hybridizations of normal samples. In this set of 20 hybridizations, ratios on six clones were outside the thresholds, leading to a presumed false positive rate of 0.4%. However, as discussed by others (40), the standard deviation of the hybridization ratios on individual clones reported in this study could be as high as 0.14, suggesting that assays using this array of 77 clones could result in 15% of assays being reported as abnormal. Thus, improvements in image analysis for arrays and procedures to recognize outliers due to hybridization or array printing artifacts will be required if genomic array-based assays are to move from their current research and proof of concept implementation into clinical diagnostic laboratories. Further, chromosome CGH has proved useful in prenatal testing for aneuploidy (44). It is likely that array-based CGH can be automated and procedures for measuring single cells (5) can be improved sufficiently to allow the technique to be applied in infertility clinics. Such applications will place even greater demands on array performance and will require rigorous procedures for data analysis to avoid false positive and negative results.

Copy number polymorphisms in the population complicate the interpretation of genomic analyses, whether by FISH based methods or array CGH. Variation in repeats and length polymorphisms have been reported in telomeres, most notably 2q and Xp/Ypter, and are found in patients as well as unaffected
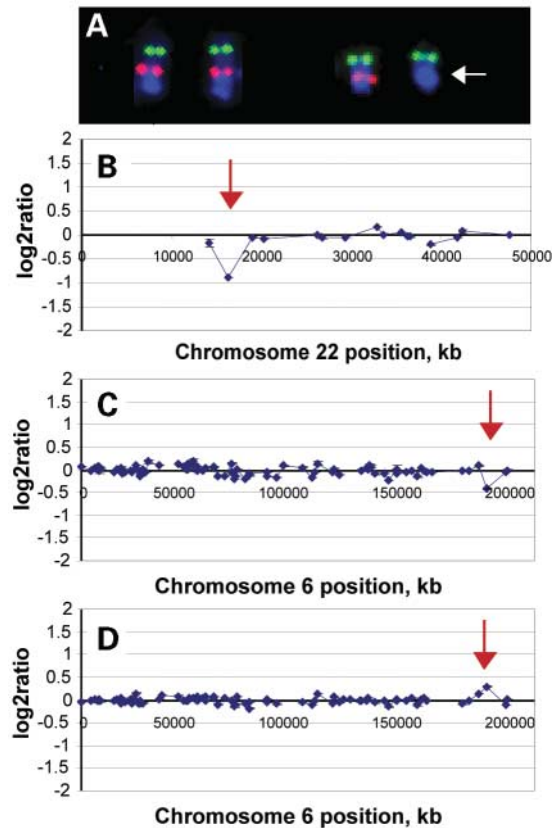

Figure 3. Detection of constitutional chromosomal aberrations and copy number polymorphisms. (**A**) Detection of deletion in the DiGeorge region by FISH. A chromosome 22 subtelomere probe (green) and the TUPLE1 probe for the DiGeorge region (red) were hybridized to metaphase chromosomes from a normal individual and an individual with the deletion. The arrow indicates the missing red FISH signal on the deleted chromosome. (**B**) Array CGH copy number profile of chromosome 22 showing deletion in the DiGeorge region (arrow). Data are plotted as in Figure 2. (**C** and **D**) Copy number polymorphism at 6q26 in two individuals (arrows). Copy number profiles are shown with decreased (C) or increased (D) copy number relative to the same reference DNA. Data are plotted as in Figure 2, and the complete data sets are available from Snijders *et al.* (2).

relatives (40,41). Other insertion/deletion germline polymorphisms have been reported that are often flanked by repeated sequences. In some cases, they may be associated with disease or they may have no phenotype (45–51). Figure 3C and D shows a common copy number variation affecting a single BAC at 6q26. The apolipoprotein(a) gene maps to this locus. It is likely that the observed copy number differences between individuals reflects variation in the length of this gene, which is highly polymorphic in the human genome due to variation in the number of copies of a 5.5 kb sequence encoding kringle repeats (52,53). Since the human genome contains many gene families and duplications that can promote insertion/deletion rearrangements (54–56), many more germline copy number polymorphisms are likely to be revealed by widespread application of array CGH. Thus, when interpreting array CGH data, it is important to bear in mind that certain copy number changes may be polymorphisms that do not affect phenotype and that some clones may be sensitive to copy number changes occurring at more than one location in the genome.

FUTURE DIRECTIONS

In this review, we have focused on microarray-based methods to assess DNA copy number by CGH. Altered DNA copy number is one of a number of mechanisms that may result in the changes in gene expression that underlie developmental abnormalities and cancer. Epigenetic changes in the genome have also been measured on arrays that were prepared from a CpG island library, thereby providing information on the methylation status of much of the genome (57–59). Genomic arrays have been used for other applications, including ascertaining evolutionary genomic changes in human and non-human primate genomes (60) and assessment of methylation status in relation to copy number aberrations in tumors (61). The use of arrays of various types, including CpG island arrays together with chromatin immunoprecipitation offers the possibility to obtain genome-wide maps of the interactions of proteins with the genome over time or in relation to disease states (62–66). It is likely that the quest to understand diverse biological questions will continue to drive development of novel uses of genomic arrays. The utility of genomic arrays will also increase as the density of coverage of the genome increases. An array of contiguous clones for chromosome 22 was reported recently (15) and a set of ~30 000 clones that provide contiguous coverage of the human genome has been assembled (Kryzwinski, submitted). Arrays assembled from such high-density clone sets allow single copy changes to be determined to ~50 kb. The introduction into the clinic of array-based assays of constitutional chromosomal aberrations and cancer genomes is also likely to occur in the near future and will require consideration of appropriate array designs, platforms and analysis procedures to meet the very stringent demands for reliable and sensitive performance.

ACKNOWLEDGEMENTS

REFERENCES

1. Kallioniemi, A., Kallioniemi, O.P., Sudar, D., Rutovitz, D., Gray, J.W., Waldman, F. and Pinkel, D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, **258**, 818–821.
2. Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K. *et al.* (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.*, **29**, 263–264.
3. Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D. and Brown, P.O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, **23**, 41–46.
4. Wessendorf, S., Schwaenen, C., Kohlhammer, H., Kienle, D., Wrobel, G., Barth, T.F., Nessling, M., Moller, P., Dohner, H., Lichter, P. *et al.* (2003) Hidden gene amplifications in aggressive B-cell non-Hodgkin lymphomas detected by microarray-based comparative genomic hybridization. *Oncogene*, **22**, 1425–1429.
5. Lage, J.M., Leamon, J.H., Pejovic, T., Hamann, S., Lacey, M., Dillon, D., Segraves, R., Vossbrinck, B., Gonzalez, A., Pinkel, D. *et al.* (2003) Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res.*, **13**, 294–307.
6. Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, **20**, 207–211.
7. Albertson, D.G., Ylstra, B., Segraves, R., Collins, C., Dairkee, S.H., Kowbel, D., Kuo, W.L., Gray, J.W. and Pinkel, D. (2000) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat. Genet.*, **25**, 144–146.
8. Cai, W.W., Mao, J.H., Chow, C.W., Damani, S., Balmain, A. and Bradley, A. (2002) Genome-wide detection of chromosomal imbalances in tumors using BAC microarrays. *Nat. Biotechnol.*, **20**, 393–396.
9. Solinas-Toldo, S., Lampel, S., Stilgenbauer, S., Nickolenko, J., Benner, A., Dohner, H., Cremer, T. and Lichter, P. (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, **20**, 399–407.
10. Klein, C.A., Schmidt-Kittler, O., Schardt, J.A., Pantel, K., Speicher, M.R. and Riethmuller, G. (1999) Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl Acad. Sci. USA*, **96**, 4494–4499.
11. Veltman, J.A., Schoenmakers, E.F., Eussen, B.H., Janssen, I., Merkx, G., van Cleef, B., van Ravenswaaij, C.M., Brunner, H.G., Smeets, D. and van Kessel, A.G. (2002) High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am. J. Hum. Genet.*, **70**, 1269–1276.
12. Hodgson, G., Hager, J.H., Volik, S., Hariono, S., Wernick, M., Moore, D., Nowak, N., Albertson, D.G., Pinkel, D., Collins, C. *et al.* (2001) Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nat. Genet.*, **29**, 459–464.
13. Telenius, H., Carter, N.P., Bebb, C.E., Nordenskjold, M., Ponder, B.A. and Tunnacliffe, A. (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics*, **13**, 718–725.
14. Fiegler, H., Carr, P., Douglas, E.J., Burford, D.C., Hunt, S., Scott, C.E., Smith, J., Vetrie, D., Gorman, P., Tomlinson, I.P. *et al.* (2003) DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer*, **36**, 361–374.
15. Buckley, P.G., Mantripragada, K.K., Benetkiewicz, M., Tapia-Paez, I., Diaz De Stahl, T., Rosenquist, M., Ali, H., Jarbo, C., De Bustos, C., Hirvela, C. *et al.* (2002) A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. *Hum. Mol. Genet.*, **11**, 3221–3229.
16. Hedenfalk, I., Ringner, M., Ben-Dor, A., Yakhini, Z., Chen, Y., Chebil, G., Ach, R., Loman, N., Olsson, H., Meltzer, P. *et al.* (2003) Molecular classification of familial non-BRCA1/BRCA2 breast cancer. *Proc. Natl Acad. Sci. USA*, **100**, 2532–2537.
17. Hyman, E., Kauraniemi, P., Hautaniemi, S., Wolf, M., Mousses, S., Rozenblum, E., Ringner, M., Sauter, G., Monni, O., Elkahloun, A. *et al.* (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.*, **62**, 6240–6245.
18. Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani, R., Botstein, D., Borresen-Dale, A.L. and Brown, P.O. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl Acad. Sci. USA*, **99**, 12963–12968.
19. Clark, J., Edwards, S., Feber, A., Flohr, P., John, M., Giddings, I., Crossland, S., Stratton, M.R., Wooster, R., Campbell, C. *et al.* (2003) Genome-wide screening for complete genetic loss in prostate cancer by comparative hybridization onto cDNA microarrays. *Oncogene*, **22**, 1247–1252.
20. Lucito, R., West, J., Reiner, A., Alexander, J., Esposito, D., Mishra, B., Powers, S., Norton, L. and Wigler, M. (2000) Detecting gene copy number fluctuations in tumor cells by microarray analysis of genomic representations. *Genome Res.*, **10**, 1726–1736.
21. Monni, O., Barlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M.L. *et al.* (2001) Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl Acad. Sci. USA*, **98**, 5711–5716.

22. Redon, R., Hussenet, T., Bour, G., Caulee, K., Jost, B., Muller, D., Abecassis, J. and du Manoir, S. (2002) Amplicon mapping and transcriptional analysis pinpoint cyclin L as a candidate oncogene in head and neck cancer. *Cancer Res.*, **62**, 6211–6217.

23. Goldstein, M., Rennert, H., Bar-Shira, A., Burstein, Y., Yaron, Y. and Orr-Urtreger, A. (2003) Combined cytogenetic and array-based comparative genomic hybridization analyses of Wilms tumors: amplification and overexpression of the multidrug resistance associated protein 1 gene (MRP1) in a metachronous tumor. *Cancer Genet. Cytogenet.*, **141**, 120–127.

24. Ishizuka, T., Tanabe, C., Sakamoto, H., Aoyagi, K., Maekawa, M., Matsukura, N., Tokunaga, A., Tajiri, T., Yoshida, T., Terada, M. *et al.* (2002) Gene amplification profiling of esophageal squamous cell carcinomas by DNA array CGH. *Biochem. Biophys. Res. Commun.*, **296**, 152–155.

25. Zhao, J., Roth, J., Bode-Lesniewska, B., Pfaltz, M., Heitz, P.U. and Komminoth, P. (2002) Combined comparative genomic hybridization and genomic microarray for detection of gene amplifications in pulmonary artery intimal sarcomas and adrenocortical tumors. *Genes Chromosomes Cancer*, **34**, 48–57.

26. Mao, X., Orchard, G., Lillington, D.M., Russell-Jones, R., Young, B.D. and Whittaker, S. (2003) Genetic alterations in primary cutaneous CD30+ anaplastic large cell lymphoma. *Genes Chromosomes Cancer*, **37**, 176–185.

27. Mao, X., Orchard, G., Lillington, D.M., Russell-Jones, R., Young, B.D. and Whittaker, S.J. (2003) Amplification and overexpression of JUNB is associated with primary cutaneous T-cell lymphomas. *Blood*, **101**, 1513–1519.

28. Sanchez-Izquierdo, D., Buchonnet, G., Siebert, R., Gascoyne, R.D., Climent, J., Karran, L., Marin, M., Blesa, D., Horsman, D., Rosenwald, A. *et al.* (2003) MALT1 is deregulated by both chromosomal translocation and amplification in B-cell non-Hodgkin lymphoma. *Blood*, **101**, 4539–4546.

29. Martinez-Climent, J.A., Alizadeh, A.A., Segraves, R., Blesa, D., Rubio-Moscardo, F., Albertson, D.G., Garcia-Conde, J., Dyer, M.J., Levy, R., Pinkel, D. *et al.* (2003) Transformation of follicular lymphoma to diffuse large cell lymphoma is associated with a heterogeneous set of DNA copy number and gene expression alterations. *Blood*, **101**, 3109–3117.

30. Wilhelm, M., Veltman, J.A., Olshen, A.B., Jain, A.N., Moore, D.H., Presti, J.C. Jr, Kovacs G. and Waldman, F.M. (2002) Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. *Cancer Res.*, **62**, 957–960.

31. Weiss, M.M., Kuipers, E.J., Postma, C., Snijders, A.M., Siccama, I., Pinkel, D., Westerga, J., Meuwissen, S.G., Albertson, D.G. and Meijer, G.A. (2003) Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene*, **22**, 1872–1879.

32. Snijders, A.N., Nowee, M.E., Fridlyand, J., Piek, J.M.J., Dorsman, J.C., Jain, A.N., Pinkel, D., van Diest, P.J., Verheijen, R.H.M. and Albertson, D.G. (2003) Genome-wide array based comparative genomic hybridization reveals genetic homogeneity and frequent copy number increases encompassing CCNE1 in Fallopian tube carcinoma. *Oncogene*, **22**, 4281–4286.

33. Snijders, A.N., Fridlyand, J., Mans, D., Segraves, R., Jain, A. N., Pinkel, D., Albertson, D. G. (2003) Shaping tumor and drug resistant genomes by instability and selection. *Oncogene*, **22**, 4370–4379.

34. Fritz, B., Schubert, F., Wrobel, G., Schwaenen, C., Wessendorf, S., Nessling, M., Korz, C., Rieker, R.J., Montgomery, K., Kucherlapati, R. *et al.* (2002) Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res.*, **62**, 2993–2998.

35. Veltman, J.A., Fridlyand, J., Pejavar, S., Olshen, A.B., Korkola, J.E., DeVries, S., Carroll, P., Kuo, W.L., Pinkel, D., Albertson, D. *et al.* (2003) Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res.*, **63**, 2872–2880.

36. Ness, G.O., Lybaek, H. and Houge, G. (2002) Usefulness of high-resolution comparative genomic hybridization (CGH) for detecting and characterizing constitutional chromosome abnormalities. *Am. J. Med. Genet.*, **113**, 125–136.

37. Rauen, K.A., Albertson, D.G., Pinkel, D. and Cotter, P.D. (2002) Additional patient with del(12)(q21.2q22): further evidence for a candidate region for cardio-facio-cutaneous syndrome? *Am. J. Med. Genet.*, **110**, 51–56.

38. Veltman, J.A., Jonkers, Y., Nuijten, I., Janssen, I., Van Der Vliet, W., Huys, E., Vermeesch, J., Van Buggenhout, G., Fryns, J.P., Admiraal, R. *et al.* (2003) Definition of a critical region on chromosome 18 for congenital aural atresia by array CGH. *Am. J. Hum. Genet.*, **72**, 1578–1584.

39. Gunn, S.R., Mohammed, M., Reveles, X.T., Viskochil, D.H., Palumbos, J.C., Johnson-Pais, T.L., Hale, D.E., Lancaster, J.L., Hardies, L.J., Boespflug-Tanguy, O. *et al.* (2003) Molecular characterization of a patient with central nervous system dysmyelination and cryptic unbalanced translocation between chromosomes 4q and 18q. *Am. J. Med. Genet.*, **120A**, 127–135.

40. Flint, J. and Knight, S. (2003) The use of telomere probes to investigate submicroscopic rearrangements associated with mental retardation. *Curr. Opin. Genet. Dev.*, **13**, 310–316.

41. van Karnebeek, C.D., Koevoets, C., Sluijter, S., Bijlsma, E.K., Smeets, D.F., Redeker, E.J., Hennekam, R.C. and Hoovers, J.M. (2002) Prospective screening for subtelomeric rearrangements in children with mental retardation of unknown aetiology: the Amsterdam experience. *J. Med. Genet.*, **39**, 546–553.

42. Knight, S.J., Lese, C.M., Precht, K.S., Kuc, J., Ning, Y., Lucas, S., Regan, R., Brenan, M., Nicod, A., Lawrie, N.M. *et al.* (2000) An optimized set of human telomere clones for studying telomere integrity and architecture. *Am. J. Hum. Genet.*, **67**, 320–332.

43. Snijders, A.M., Nowak, N., Segraves, R., Blackwood, S., Brown, N., Conroy, J., Hamilton, G., Hindle, A.K., Huey, B., Kimura, K. *et al.* (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.*, **29**, 263–264.

44. Wells, D. and Levy, B. (2003) Cytogenetics in reproductive medicine: the contribution of comparative genomic hybridization (CGH). *Bioessays*, **25**, 289–300.

45. Ryu, B., Song, J., Sohn, T., Hruban, R.H. and Kern, S.E. (2001) Frequent germline deletion polymorphism of chromosomal region 8p12-p21 identified as a recurrent homozygous deletion in human tumors. *Genomics*, **72**, 108–112.

46. Giglio, S., Broman, K.W., Matsumoto, N., Calvari, V., Gimelli, G., Neumann, T., Ohashi, H., Voullaire, L., Larizza, D., Giorda, R. *et al.* (2001) Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. *Am. J. Hum. Genet.*, **68**, 874–883.

47. Lin, H., Pizer, E.S. and Morin, P.J. (2000) A frequent deletion polymorphism on chromosome 22q13 identified by representational difference analysis of ovarian cancer. *Genomics*, **69**, 391–394.

48. Rahmutula, D., Nakayama, T., Soma, M., Takahashi, Y., Uwabo, J., Sato, M., Izumi, Y., Kanmatsuse, K. and Ozawa, Y. (2000) An insertion/deletion polymorphism in intron 18 of the type B human natriuretic peptide receptor gene is not associated with cerebral infarction. *Hypertens. Res.*, **23**, 173–176.

49. Rieder, M.J., Taylor, S.L., Clark, A.G. and Nickerson, D.A. (1999) Sequence ariation in the human angiotensin converting enzyme. *Nat. Genet.*, **22**, 59–62.

50. Schunkert, H. (1997) Polymorphism of the angiotensin-converting enzyme gene and cardiovascular disease. *J. Mol. Med.*, **75**, 867–875.

51. Su, G.H., Bansal, R., Murphy, K.M., Montgomery, E., Yeo, C.J., Hruban, R.H. and Kern, S.E. (2001) ACVR1B (ALK4, activin receptor type 1B) gene mutations in pancreatic carcinoma. *Proc. Natl Acad. Sci. USA*, **98**, 3254–3257.

52. Lackner, C., Cohen, J.C. and Hobbs, H.H. (1993) Molecular definition of the extreme size polymorphism in apolipoprotein(a). *Hum. Mol. Genet.*, **2**, 933–940.

53. Kamboh, M.I., Ferrell, R.E. and Kottke, B.A. (1991) Expressed hypervariable polymorphism of apolipoprotein (a). *Am. J. Hum. Genet.*, **49**, 1063–1074.

54. Ji, Y., Eichler, E.E., Schwartz, S. and Nicholls, R.D. (2000) Structure of chromosomal duplicons and their role in mediating human genomic disorders. *Genome Res.*, **10**, 597–610.

55. Cheung, J., Estivill, X., Khaja, R., MacDonald, J.R., Lau, K., Tsui, L.C. and Scherer, S.W. (2003) Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence. *Genome Biol.*, **4**, R25.

56. Lupski, J.R. (1998) Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.*, **14**, 417–422.

57. Yan, P.S., Wei, S.H. and Huang, T.H. (2002) Differential methylation hybridization using CpG island arrays. *Meth. Mol. Biol.*, **200**, 87–100.

58. Shi, H., Maier, S., Nimmrich, I., Yan, P.S., Caldwell, C.W., Olek, A. and Huang, T.H. (2003) Oligonucleotide-based microarray for DNA methylation analysis: principles and applications. *J. Cell. Biochem.*, **88**, 138–143.

59. Yan, P.S., Chen, C.M., Shi, H., Rahmatpanah, F., Wei, S.H., Caldwell, C.W. and Huang, T.H. (2001) Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res.*, **61**, 8375–8380.

60. Locke, D.P., Segraves, R., Carbone, L., Archidiacono, N., Albertson, D.G., Pinkel, D. and Eichler, E.E. (2003) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.*, **13**, 347–357.

61. Zardo, G., Tiirikainen, M.I., Hong, C., Misra, A., Feuerstein, B.G., Volik, S., Collins, C.C., Lamborn, K.R., Bollen, A., Pinkel, D. *et al.* (2002) Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat. Genet.*, **32**, 453–458.

62. Wells, J., Yan, P.S., Cechvala, M., Huang, T. and Farnham, P.J. (2003) Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene*, **22**, 1445–1460.

63. Horak, C.E. and Snyder, M. (2002) ChIP-chip: a genomic approach for identifying transcription factor binding sites. *Meth. Enzymol.*, **350**, 469–483.

64. Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M. and Snyder, M. (2002) GATA-1 binding sites mapped in the beta-globin locus by using mammalian chIp-chip analysis. *Proc. Natl Acad. Sci. USA*, **99**, 2924–2929.

65. Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature*, **409**, 533–538.

66. Fontemaggi, G., Kela, I., Amariglio, N., Rechavi, G., Krishnamurthy, J., Strano, S., Sacchi, A., Givol, D. and Blandino, G. (2002) Identification of direct p73 target genes combining DNA microarray and chromatin immunoprecipitation analyses. *J. Biol. Chem.*, **277**, 43359–43368.

67. Albertson, D.G., Collins, C., McCormick, F. and Gray, J. W. (2003) Chromosome Aberrations in Solid Tumors. *Nat. Genet.*, **34**, 369–376.