

Genomic microarrays in the spotlight

Kiran K. Mantripragada*, Patrick G. Buckley*, Teresita Diaz de Ståhl and Jan P. Dumanski

Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, 751 85 Uppsala, Sweden

Microarray-based comparative genomic hybridization (array-CGH) has emerged as a revolutionary platform, enabling the high-resolution detection of DNA copy number aberrations. In this article we outline the use and limitations of genomic clones, cDNA clones and PCR products as targets for genomic microarray construction. Furthermore, the applications and future aspects of these arrays for DNA copy number analysis in research and diagnostics, epigenetic profiling and gene annotation are discussed. These recent developments of genomic microarrays mark only the beginning of a new generation of high-resolution and high-throughput tools for genetic analysis.

The application of comparative genomic hybridization (CGH) to metaphase spreads enabled genome-wide analysis of gross DNA copy number imbalances [1]. This conceptually novel approach uses differentially labeled 'test' and 'reference' DNA, which are competitively hybridized to normal metaphase chromosomes on a glass slide. The ratio of the fluorescence intensities detected is indicative of the relative DNA copy number in test versus reference DNA. However, the main disadvantage of metaphase-CGH is its low resolution: values documented in the literature are 5–10 Mb for deletions and 2 Mb for amplifications. This limitation was resolved with the advent of microarray-based CGH. First described in 1997, matrix-CGH (also known as array-CGH) paved the way for higher resolution detection of DNA copy number aberrations [2]. Array-CGH is based on the same principles as metaphase-CGH, except that the targets are mapped genomic clones instead of whole chromosomes (Figure 1). This approach introduced a new dimension to DNA copy number detection in terms of resolution and specificity of analysis. In 1998, Pinkel *et al.* examined the sensitivity and quantitative capability of array-CGH for gene dosage measurements using an arrayed subset of genomic clones that demonstrated the usefulness of this technique in detecting cancer-associated DNA copy number aberrations [3]. The first genome-wide array-CGH analysis of DNA losses and gains used immobilized cDNA clones as hybridization targets. This study used an array containing 3360 human cDNAs to map gene amplifications and deletions in breast cancer [4]. In 2001, Snijders *et al.* described a genome-wide DNA microarray that consisted of 2400 BACs distributed across the genome [5]. Recently, we

reported the first microarray contiguously covering a human chromosome with an average resolution of 75 kb [6].

This article will focus on the production, applications and future aspects of genomic microarrays. The convention used throughout this review is to define the DNA immobilized on a solid support as the 'target' and to define a genome (or a fraction of the genome) applied to an array during a hybridization experiment as the 'probe'.

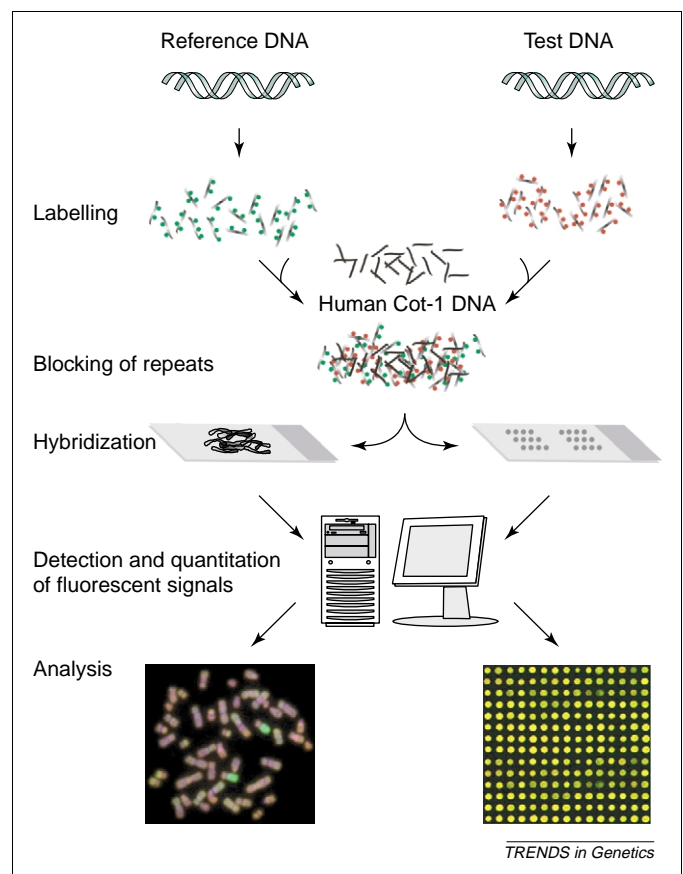


Figure 1. Comparative genomic hybridization (CGH) on metaphase spreads and arrays enables genome-wide analysis of DNA sequence copy number in a single experiment. Test and reference genomic DNA are differentially labeled, mixed, co-precipitated with unlabeled Cot-1 DNA to enable the blocking of repetitive sequences and dissolved in hybridization solution. A sufficient amount of Cot-1 DNA is important to obtain adequate suppression of the repetitive sequences. The hybridization solution containing the probe is deposited on the slide containing either metaphase or array targets and hybridized for 16–72 h. After hybridization, the slide is washed and fluorescence signals are detected. Using image-processing software, chromosomal regions with an abnormal test:reference ratio, that is regions with a loss or gain of DNA sequences, can be detected. The use of metaphase chromosomes is however limited by its low resolution. The development of array-CGH has eliminated the need to obtain metaphase spreads and has dramatically increased the resolution of the technique.

* These authors contributed equally to this article.

Corresponding author: Jan P. Dumanski (jan.dumanski@genpat.uu.se).

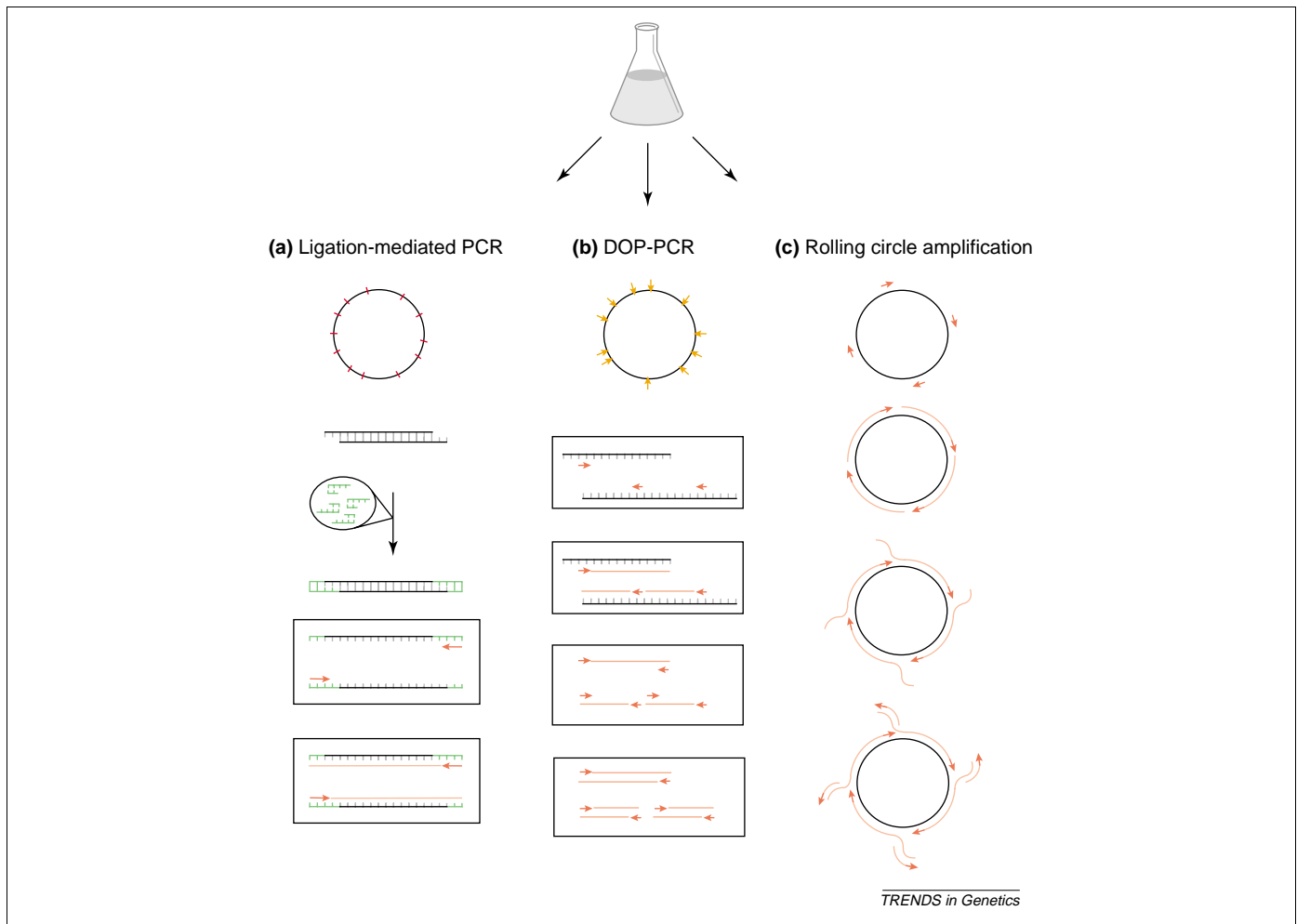


Figure 2. DNA amplification strategies for the production of genomic microarrays. The three strategies outlined amplify genomic template DNA that is derived from bacterially cultured and purified vector DNA. **(a)** Ligation-mediated PCR. Genomic DNA is initially digested to generate fragments, ranging from 200 to 2000 bp. An oligonucleotide adaptor (shown in green) is then ligated, which generates a universal priming site for all DNA fragments. Primers (depicted by the red arrows) that are complementary to the adaptor are used to amplify the fragments by PCR. **(b)** Degenerate oligonucleotide-primed PCR (DOP-PCR) uses partially degenerate primers to amplify genomic DNA, using a low annealing temperature initially in the PCR reaction. DOP-PCR primers are designed to preferentially amplify human DNA in the presence of contaminating *Escherichia coli* DNA. **(c)** Rolling circle amplification uses exonuclease-resistant random primers, coupled with ϕ -29 polymerase, to amplify DNA in an isothermal, strand-displacing reaction.

Strategies for array construction

Genomic clones: use and limitations

The current methodological approach for the construction of genomic arrays is dominated by the use of genomic clones (BACs, PACs and cosmids) as targets for array hybridization. These vectors are propagated in bacteria, purified and then spotted onto a glass slide. This approach is labor intensive because PACs and BACs are single-copy vectors and yield a low amount of DNA. To circumvent this difficulty, several strategies have been developed that amplify small amounts of cultured and purified clone DNA for the production of thousands of slides. Figure 2 outlines these approaches.

Ligation-mediated PCR was the first reported technique for the amplification of clones used in array-CGH [5,7]. This method is based on the digestion of DNA with a commonly cutting restriction endonuclease that results in fragments ranging from 200 to 2000 bp. A universal oligonucleotide adaptor is then ligated, which serves as the priming site for PCR amplification. The second approach uses degenerate oligonucleotide-primed polymerase chain

reaction (DOP-PCR) to randomly amplify BAC DNA [8–11]. This technique has been refined recently by selecting three degenerate primers that preferentially amplify human DNA compared with contaminating *Escherichia coli* DNA. The pooling of these three individual DOP-PCR amplifications resulted in a more representative sequence when compared with the method reported originally [11]. Finally, rolling circle amplification has been shown to amplify efficiently from minute amounts of DNA. Enzymes, such as ϕ -29 DNA polymerase from the ϕ -29 phage of *Bacillus subtilis*, can perform strand displacing, proof-reading and isothermal amplifications of circular and large linear DNA molecules. Several rolling circle amplification techniques have been reported that amplify clones for array production [6] and patient DNA for hybridization [12–14]. Although these methods decrease the time and labor of DNA production, while increasing the yield, they all rely on an initial preparation of template DNA from the respective genomic clone.

The genomic clone-based approach has proven to be successful in a large number of applications and it is likely that it will continue to dominate in the coming years

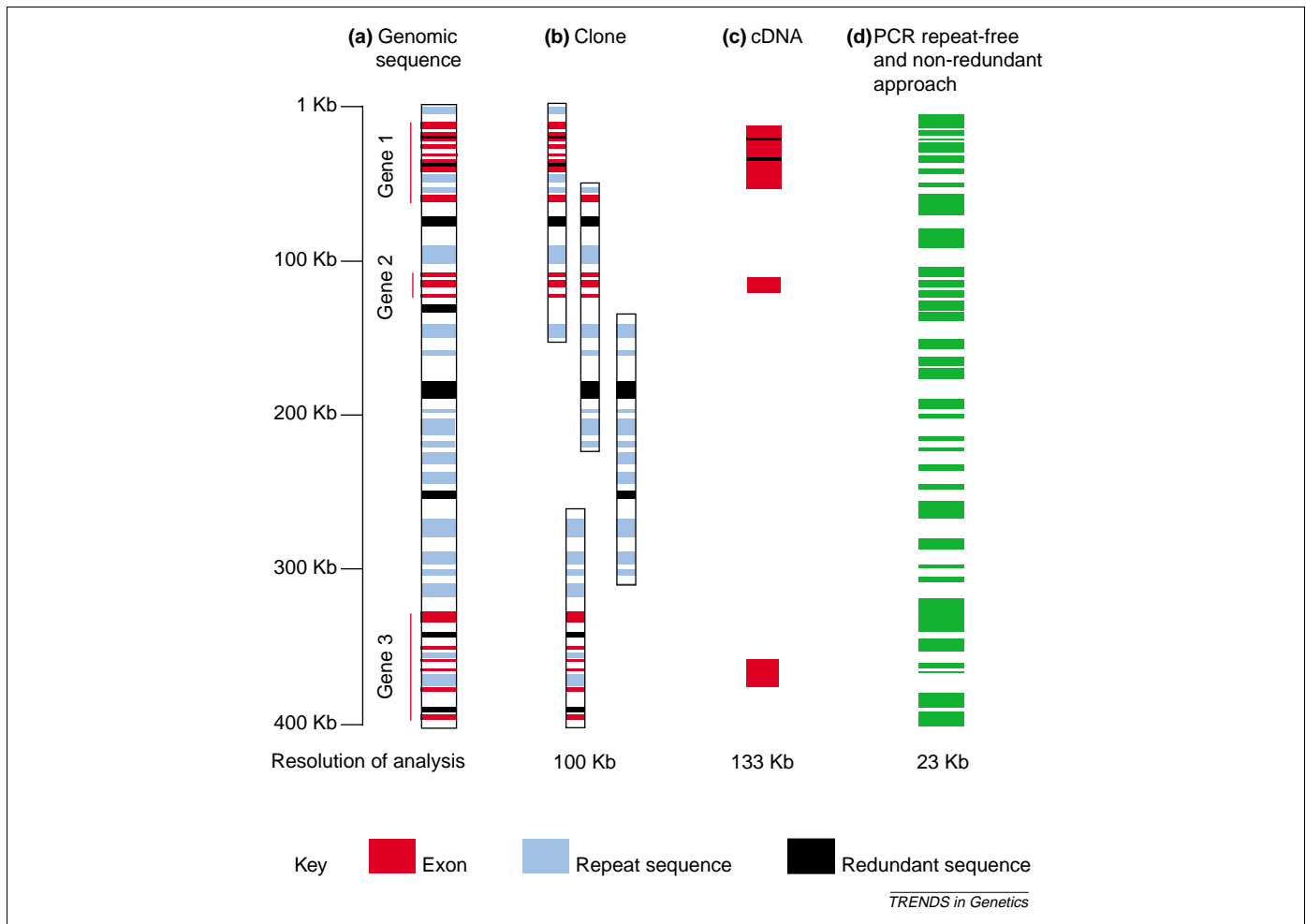


Figure 3. Strategies for the construction of genomic arrays. **(a)** A 400-kb genomic sequence is presented schematically. Features present in this sequence are highlighted. Gene exons and common repeat sequences are represented by red and blue, respectively, whereas other redundant sequences [e.g. low copy repeats (LCRs), pseudogenes, paralogous genes and gene segments] are denoted in black. **(b)** Description of the clone-based (cosmid, BAC or PAC) approach for contiguous coverage of genomic sequence. This strategy uses publicly available clones for the construction of microarray targets. However, these clones contain all the features that are present in the genomic sequence and thus fail to perform when the DNA sequence increases in complexity. The highest reported resolution of contiguously covering genomic segments using this method is ~75 kb [6]. **(c)** The use of cDNA clones for the construction of microarray targets. This method uses cDNA clones for the analysis of genomic imbalances. **(d)** PCR repeat-free and non-redundant approach for microarray construction. This method relies on the bioinformatic selection of repeat-free and non-redundant sequences. These regions are then amplified by PCR and spotted onto glass slides to represent unique sequences that correspond to a particular genomic sequence. The highest reported resolution using this method is ~23 kb [16]. Reproduced with permission from Ref. [6].

(Figure 3). This procedure, however, has important limitations. The first restriction is the resolution of analysis, which is automatically determined by the size of inserts in genomic clones (~40 kb for a cosmid and ~100 kb for a BAC) and the map distance between the targets printed in each spot. Furthermore, the spotting of genomic clones does not permit pre-selection of the sequence present in these clones. This results in an array that contains all common repeats [e.g. *Alu* and long interspersed nuclear elements (LINEs)], other redundant sequences such as low copy repeats [LCRs (also known as segmental duplications)] centromeric and telomeric repeats and segments of extensive sequence similarity (pseudogenes or paralogous genes) [15]. These sequences lead to complications in data interpretation in array-CGH experiments [6,11,16].

Use of cDNA arrays for DNA copy number profiling

cDNA microarrays, first described in 1995 [17], are routinely used to characterize variation in gene expression. In 1999, Pollack *et al.* [4] reported the application of cDNA

microarrays for the analysis of DNA copy number (Figure 3). There are several advantages in employing cDNAs as targets for array-CGH hybridizations. The widespread availability of cDNA clone-sets enables the large-scale production of microarrays. In addition, this platform presents the possibility of performing parallel analysis of changes in DNA copy number and expression levels using the same set of genes. Recently, a parallel genome-wide expression and DNA copy number analysis across 6691 mapped human genes was performed in breast cancer [18]. This study showed that 62% of highly amplified genes demonstrated moderately or highly elevated expression. Such cDNA-based microarray analyses have identified other genes that are important in the pathogenesis of several disorders [19–23].

However, there are also limitations involved in the use of cDNA clones in array-CGH. First, cDNA microarray analysis enables only the detection of aberrations in known genes. Although gene-rich regions are analyzed at a high resolution, the so called 'gene deserts' (or regions that are not well studied with regard to transcriptionally

active genes) are completely excluded from analysis. This results in an uneven distribution of measurement points across the genome, which is an important concern in view of the recent reports on the transcriptional activity of the human genome [24,25]. These studies suggest that the transcriptional activity of the human genome is 2–10 times higher than the number of currently annotated genes. Second, promoters, introns and intergenic sequences involved in gene regulation are not represented on cDNA arrays, which limit their application, especially in epigenetic studies. Third, the level of sensitivity and specificity for detection of single-copy gene deletions and low-copy number gains using cDNA microarrays requires further improvements [4,19,22]. Each individual cDNA array element provides a low level of sensitivity and specificity for the detection of single-copy gene deletions [4]. Finally, some cDNA clones contain some redundant sequences because of extensive sequence similarities between paralogous genes, which complicate the interpretation of array CGH data.

Repeat-free and non-redundant approach

As described previously, the spotting of genomic or cDNA clones as targets for array-CGH hybridizations carries inherent limitations, particularly in relation to the resolution of analysis and applicability of the method to 'difficult' genomic regions that are particularly rich in common repeats and/or contain segmental duplications. We have recently addressed these innate problems related to genomic- and cDNA-clone-based approaches for array-CGH in two separate studies. The first study used a cosmid-based array covering 140 kb across neurofibromin 2 (*NF2*), a tumor suppressor gene located on chromosome 22q12. We observed that a high content (up to 80%) of common repeats within the first half of this locus led to unreliable data, despite the inclusion of high quantities of human Cot-1 DNA in the hybridization solution [16]. The second study focused on the DiGeorge syndrome region in chromosome 22q11, where problems of false-negative results (e.g. deleted regions scored as non-deleted) during array-CGH analysis were related to both the presence of LCRs and the high content of common repeats [26]. To circumvent these problems, we designed and validated a PCR-based, strictly sequence defined, repeat-free strategy for the generation of microarray targets (Figure 3). Using this approach, repeat-free and non-redundant sequences were amplified by PCR from genomic clones and were spotted in pools as targets on a glass slide. The resolution of analysis depends on the pool size, in terms of genomic DNA covered by the PCR products. We successfully detected disease-causing deletions at a resolution of 23 kb and 15 kb in the *NF2* gene and the DiGeorge syndrome region, respectively. The *NF2* gene-specific array for diagnostics of intragenic deletions is the first array of this type [16]. The repeat-free approach has advantages in research and diagnostic settings and can be applied to any region of the genome for DNA copy number analysis, regardless of the content of the redundant sequence. In addition, this strategy provides the possibility of further increasing the resolution of analysis. It is likely that the limit of 15–20 kb per measurement point will be decreased dramatically in the near future.

At present, the main disadvantage of this approach is that it can only be applied to sequenced genomic segments. Further drawbacks include the high cost of primers and the labor-intensive procedures for the production of thousands of PCR fragments, if the latter is performed without automation. Another drawback is the lack of efficient bioinformatic tools, which would enable fully automated design of larger (on the chromosomal scale) arrays. Towards this goal, we have recently developed software (named 'Sequence Allocator'), which combines *RepeatMasker* (<http://repeatmasker.genome.washington.edu>), *blastn* (www.ncbi.nlm.nih.gov) and *Primer3* (http://www.broad.mit.edu/genome_software/other/primer3.html) programs into a fully automated bioinformatic protocol for the identification of non-redundant sequences and design of PCR primers (U. Menzel *et al.*, unpublished). A demo version is available from the website of the authors upon request (<http://puffer.genpat.uu.se>).

Applications of genomic microarrays

Array-based detection of DNA copy number variation in research

Array-CGH is currently the most powerful method to detect and localize simultaneously loss or gain of genetic material. High-resolution mapping of such specific regions is of crucial importance for the subsequent discovery of the disease-associated gene(s). Clone-based arrays covering the whole genome, with an average resolution of 1.3 Mb, have been applied successfully to gain insights into the pathogenetic mechanisms of cancer development and evolutionary studies [5,27–34]. We have constructed an array that provides full coverage of human chromosome 22 with an average resolution of 75 kb [6]. The usefulness of the array is demonstrated by the detection of a heterozygous deletion spanning most of the chromosome plus a homozygous deletion encompassing the gene encoding 14-3-3-eta, while scanning 34 Mb of chromosome 22q in a limited number of glioblastomas [6] (Figure 4). Recently, a clone-set that contains ~32 000 BACs and provides contiguous coverage of the human genome has been assembled (M. Krzywinski *et al.*, unpublished; <http://www.bcgsc.ca/lab/mapping/bacarray/human/>). Future arrays containing such a high density of clones will enable genome-wide analysis with an average resolution of 100 kb. This will have a considerable impact on research directed towards the characterization of human disease-related genes and basic genetic research.

Although array-CGH has been used mostly in cancer research to date, its application in the study of congenital and acquired genetic diseases is increasing [10,35]. The usefulness of array-CGH has been shown by the identification of deletions in patients with DiGeorge syndrome [6,26] and by the detection of terminal and interstitial deletions, derivative chromosomes and complex rearrangements in patients with mental retardation [36].

Genomic arrays as gene annotation tools

The use of genomic microarrays for the analysis of transcriptional activity at the chromosomal level has provided interesting insights into the number of potential

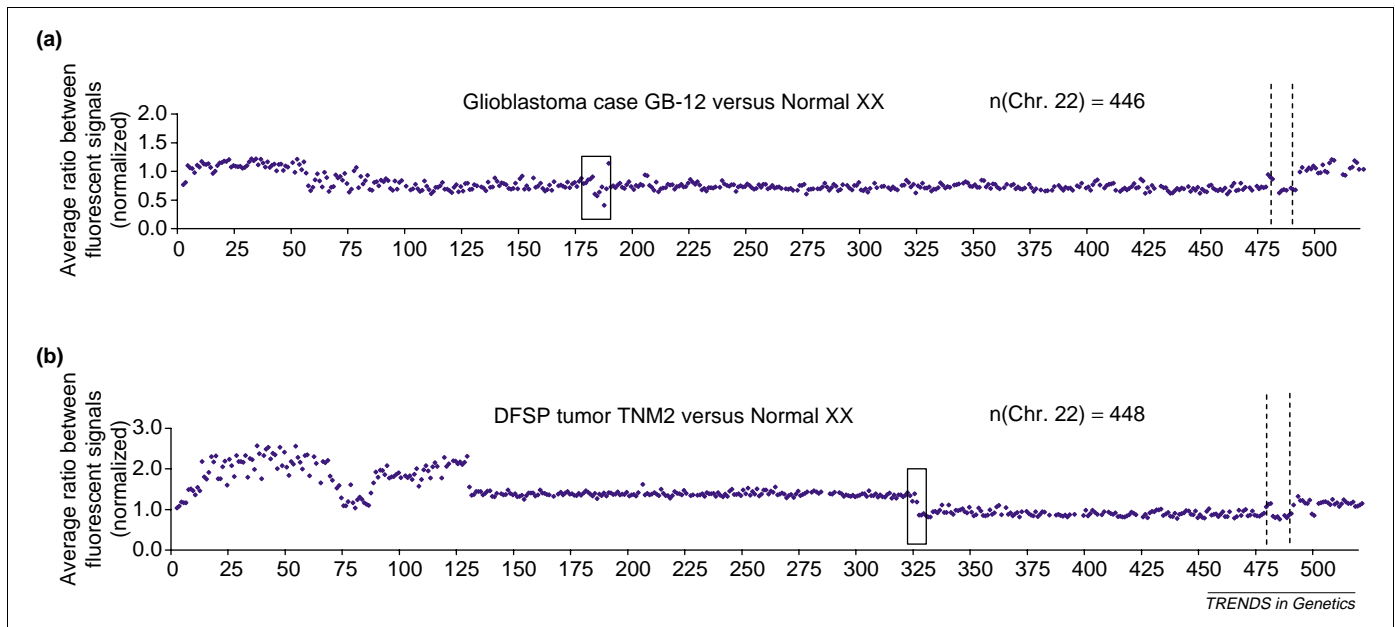


Figure 4. Detection of a homozygous deletion and an unbalanced translocation using comparative genomic hybridization arrays (array-CGH) [6]. The x-axis displays 480 chromosome 22 loci (ID1–480), ordered from the centromere (left-hand side) to the telomere (right-hand side). The dots between the broken vertical lines indicate control loci from chromosome X (nine clones). The dots plotted after the chromosome X controls are derived from chromosomes other than X and 22 (31 clones). The $n(\text{Chr. 22})$ -value denotes the number of chromosome 22q loci that were scored in a single experiment. Chromosomal X control clones are particularly useful in hybridizations between DNA from males and females because they are indicative of the haploid level of the fluorescence ratio. The average fluorescence ratios from non-chromosome 22 controls indicate the diploid level of DNA copy number. Each dot on a chart represents the normalized, average ratio between fluorescent signals for each locus (y-axis), which are derived from 3–4 independent replica spots on the array, in a single hybridization experiment. (a) Detection of a novel, homozygous deletion in glioblastoma case GB-12 from a male patient. The homozygously deleted cosmid (highlighted in the rectangle), containing the 14-3-3 eta gene displayed a normalized CGH ratio of 0.35. The non-zero value for the ratio in the homozygously deleted region might be due to a normal tissue component in the tumor, inefficient blocking of repetitive sequences, autofluorescence of the target clones or that cosmid cN44A4 is larger than the deletion. Both copies of chromosome 22, situated between the peri-centromeric region (ID1) and locus ID55, are retained in the tumor. The large terminal, heterozygous deletion encompasses ID56–480 and displays fluorescence ratios consistent with the level of chromosome X controls (haploid level). (b) Analysis of an unbalanced $t(17;22)(q22;q13)$ translocation in dermatofibrosarcoma protuberans (DFSP) tumor DNA from a male patient TNM2. The rectangle outlines the translocation breakpoint at cosmid cN10C3, containing the *PDGFB* gene. The region distal to the *PDGFB* gene on 22q displayed fluorescence ratios consistent with one DNA copy level, as a result of the loss of *der(17)* chromosome. We also identified two distinct amplifications between ID5 and ID128. Reproduced with permission from Ref. [6].

coding sequences in the human genome [24,25]. Using an Affimetrix platform, Kapranov *et al.* generated 25-nucleotide oligomers spaced every 35 bp across human chromosomes 21 and 22 [24]. Another study produced an array representing the majority of unique sequence of chromosome 22 to assess its transcriptional activity [25]. These studies identified 2–10 times more transcripts when compared with the annotated genes currently in the databases. The main advantage of these genomic array-based approaches over mainstream cDNA arrays is that they identify previously uncharacterized regions with high transcriptional activity and true ‘gene deserts’.

Use of array-CGH in diagnostics

Genetic aberrations at the DNA copy number level, such as deletions, amplifications and unbalanced translocations, are among the main pathogenetic mechanisms that underlie human genetic disorders (<http://www.ncbi.nlm.nih.gov/omim/>). Analysis of chromosomal aberrations is particularly important in cancer, where amplifications of oncogenes and/or deletion of tumor suppressor genes are involved in the multi-step process of cancer development [37]. Therefore, rapid and accurate identification of such genetic imbalances could improve diagnostics, which might lead to better decision-making regarding the choice of available treatments. However, despite the fact that the disease-causing genetic imbalances are important, the routine approaches for detection of such imbalances are

underdeveloped. The need for improved methodology for DNA copy number analysis is particularly apparent when compared with a variety of methods available for the accurate detection of point mutations.

Arrays focused on selected segments of the genome that are known to be frequently associated with specific diseases have been applied successfully [16,20,38–42]. One such example is neurofibromatosis type 2 (NF2) (OMIM# 101 000), which is an autosomal dominant disorder caused by the bi-allelic inactivation of *NF2* on chromosome 22q12. Affected individuals usually develop bilateral schwannomas of the eighth cranial nerve but also can develop other tumors. Inactivating deletions in *NF2* occur in 20–30% of NF2 patients. A high-resolution diagnostic microarray for profiling DNA copy number alterations in *NF2* has been developed recently [16]. Because array-CGH is used predominantly in cancer-related research, it is challenged by the presence of heterogeneity in the tumor tissue. This means that the tumor-specific DNA probe used in hybridizations often contains variable amounts of DNA derived from contaminating normal cells. It has been shown recently that array-CGH has the power to discriminate between one and two DNA copies in a heterogeneous mixture of tumor (containing an aberration) and normal cells [33]. However, the limit of this technology in this respect has not yet been tested.

Array-CGH has also been shown to be useful in the analysis of translocations. We have reported the detection

of an unbalanced tumor translocation using array-CGH in dermatofibrosarcoma protuberans (DFSP) [6]. The translocation breakpoint was detected in a single cosmid (after analysis of 34.7 Mb of chromosome 22q) that was truncated as a consequence of the translocation (Figure 4b). An important drawback of conventional array-CGH is the inability to detect balanced rearrangements, which do not lead to change in the DNA copy number. In the case of balanced translocations, this limitation can be circumvented by flow-sorting of the derivative chromosomes followed by DOP-PCR amplification and hybridization to genomic arrays. This approach has been reported recently using a genomic array covering three human chromosomes, with a resolution of ~1 Mb [43]. In the near future, array-CGH has the potential to be the dominant technique in diagnosis of chromosomal imbalances, owing to its high-throughput and high-resolution of analysis and the possibility of automation.

Genetic variation in human populations at the level of micro-deletions and micro-gains

The availability of the draft sequence of the human genome enables the investigation of the role of sequence variation as the basis of human disease [44,45]. Studies of single nucleotide polymorphisms (SNPs), the most common type of variation, offer the possibility of identifying disease loci, and this is currently a dominating theme in genetic research. The SNP Consortium has developed a map of human genome variation, which contains ~2.2 million polymorphisms (www.ncbi.nlm.nih.gov/SNP). However, SNPs are not the only type of variation present in the human genome. The extent of germline and somatic polymorphisms at the level of micro-deletions and micro-gains (from 100 bp to 4 Mb and not detectable by standard cytogenetic methods) in human populations is practically unknown. This is mainly due to the lack of reliable high-resolution and high-throughput methodology. It is only recently that array-CGH has emerged as a platform with promise in this direction. Micro-deletions and micro-gains might be at least as important as SNPs for the detection of predisposing factors for disease. It should be stressed that most mutation scanning methods are based on PCR with subsequent analysis being qualitative rather than quantitative. Without quantification, heterozygous deletions and duplications might be ignored and could therefore be largely underestimated. This could impact on the outcome of SNP studies, where large-scale genotyping might miss heterozygous deletions, which would be scored as homozygous by SNP analysis. Therefore, comprehensive mutation analysis should include scanning for insertions, deletions and duplications [46].

LCRs present in the human genome have been shown to predispose to chromosomal rearrangements, often resulting in deletions and duplications. The importance of LCRs in the generation of micro-deletions and micro-gains and the relationship of LCRs to human disease is increasing since they were found to be involved in chromosome 22-associated conditions such as the DiGeorge syndrome [47]. Genetic diseases involving other chromosomes have also been shown to be associated with LCR-mediated interstitial duplications or deletions [47–49]. In the

future, the availability of human genomic arrays with an average resolution of 1–3 kb will enable the evaluation of frequency and phenotypic importance of micro-deletions and micro-gains in human populations.

Epigenetic modifications and genomic microarrays

The term epigenetics can be defined as heritable or acquired changes in gene expression that are not associated with changes in DNA sequence. The best-known epigenetic modification is DNA methylation, which tags cytosine in CpG dinucleotides with a methyl group. DNA methylation of CpG islands within the promoter and/or the first exon of genes, is generally associated with the silencing of gene expression. Chemical modification of histones can also function as an epigenetic mechanism and genes associated with acetylated histones are usually turned 'on' [50–55]. Aberrant DNA methylation within CpG islands is among the first and most common alteration leading to silencing or overexpression of genes in cancer [50,56] and is therefore potentially a promising tool for early diagnosis in cancer [55]. Conventional methods used to determine methylation status are unable to evaluate methylation changes on a genome-wide scale.

The incorporation of epigenetic analysis methods into microarray technology can yield quantitative and qualitative information of DNA methylation throughout the genome [57]. A few different approaches have been reported. In one approach, restriction landmark genomic scanning-based methylation analysis, in conjunction with array-CGH, has been applied in the identification of promoter-mediated methylation silencing of genes in cancer [58]. Another approach includes the use of gene expression microarrays to identify candidate genes that are silenced by methylation. Detection of upregulated transcripts after demethylation reflects genes that are most likely to be silenced by promoter hypermethylation [59–62]. In the third approach, specialized arrays consisting of CpG island clones have been used in combination with a method called differential methylation hybridization. Linker-ligated genomic DNA is digested with a methylation-sensitive restriction enzyme, amplified by PCR and hybridized to the array. Hypermethylated CpG islands are protected from methylation-sensitive cleavage and therefore can be amplified by PCR, producing array hybridization signals [63–65]. A simpler variant of the latter technique has been used to analyze the methylation of genes in *Arabidopsis thaliana* [66]. Finally, methylation-specific oligonucleotide-based microarrays have been used to analyze DNA methylation [67,68]. After bisulfite modification, DNA is amplified by PCR and hybridized to the oligonucleotide array. This analysis discriminates bisulfite-converted (TpG) from methylation-protected (CpG) dinucleotides, at specific nucleotide positions.

DNA microarrays can also be used to map the location of chemical modifications of chromatin and to identify chromosome targets of proteins involved in DNA binding or chromatin remodeling [57]. The 'ChIP-to-chip' technique, which combines chromatin immuno-precipitation (ChIP) with hybridization to DNA arrays, has proven to be a powerful method in elucidating the interaction sites of DNA-binding proteins throughout the genome [69,70].

Protein–DNA interactions are captured *in vivo* by crosslinking DNA-binding proteins to their binding sites. The DNA is fragmented and co-immunoprecipitated with the protein of interest using a specific antibody. Subsequently, isolated DNA fragments are amplified by non-specific PCR, labeled and hybridized to the microarray to reveal their genomic location. However, this technique depends on the linear amplification of co-precipitated DNA. It is likely that in the future, DNA extracted from a single reaction will be used directly in a hybridization probe without a pre-amplification step.

Future developments

The development of genomic microarrays, outlined in this review, marks only the beginning of a new generation of high-resolution and high-throughput tools for genetic analysis. There is room for extensive improvements in existing genomic microarray protocols on several levels. In the future, it is likely that long oligonucleotide-based arrays (comprising 50–100mers) will largely replace the BAC- and PCR-based microarrays. The primary advantage of this strategy is that it will enable a significant improvement in resolution and coverage at a reasonable cost. Recently, Lucito *et al.* [71] reported the use of 85 000 oligonucleotides (70mers) in the DNA copy number analysis of cancer-related conditions by employing genomic representations as hybridization probes. Signal-to-noise ratio during array hybridization experiments is the most important factor that determines the reliability of microarray data. It would therefore be desirable to develop more efficient fluorescent dyes and improved slide-surface chemistries. In addition, the application of microfluidics has enormous potential to improve the kinetics of hybridization by circulating a small volume of probe DNA [72]. Furthermore, the density of spotting at which most of the commercially available printing robots perform is in the range of 150–300 μm center-to-center distance. Optimal printing protocols should permit a two- to fourfold increase compared with current values, enabling the printing of a considerably higher number of features per cm^2 on the array. This would decrease the amount of often precious patient material used in hybridization and decrease experimental costs. It would also enable spotting of in excess of a million features per standard microarray slide. A further area of development is the application of bioinformatics and robotics in the initial design and automated production of high-resolution arrays, which target large segments of the genome. The integration of other methods, such as laser capture microdissection, with protocols for whole genome amplification and subsequent hybridization to high-resolution genomic arrays will permit genome-wide profiling of a single cell. We foresee that future microarrays will enable integrated genome-wide analysis of expression, epigenetic and DNA copy number at an average resolution of 1–3 kb, in a single experiment.

Acknowledgements

We thank Ian Dunham, Lore Zech, Anna Wedell, Nils-Erik Heldin and Magnus Nordenskjöld for helpful comments. We also thank the anonymous referees for their constructive criticism. This work was supported by grants from the U.S. Army Medical Research and Materiel

Command, award no. DAMD17-00-1-0536, the Swedish Cancer Foundation, the Swedish Research Council and Uppsala University.

References

- Kallioniemi, A. *et al.* (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258, 818–821
- Solinas-Toldo, S. *et al.* (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 20, 399–407
- Pinkel, D. *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* 20, 207–211
- Pollack, J.R. *et al.* (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* 23, 41–46
- Snijders, A.M. *et al.* (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* 29, 263–264
- Buckley, P.G. *et al.* (2002) A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. *Hum. Mol. Genet.* 11, 3221–3229
- Klein, C.A. *et al.* (1999) Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc. Natl. Acad. Sci. U. S. A.* 96, 4494–4499
- Telenius, H. *et al.* (1992) Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow sorted chromosomes. *Genes Chromosomes Cancer* 4, 257–263
- Telenius, H. *et al.* (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by single degenerate primer. *Genomics* 13, 718–725
- Veltman, J.A. *et al.* (2002) High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am. J. Hum. Genet.* 70, 1269–1276
- Fiegler, H. *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
- Dean, F.B. *et al.* (2002) Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5261–5266
- Lage, J.M. *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
- Hosono, S. *et al.* (2003) Unbiased whole-genome amplification directly from clinical samples. *Genome Res.* 13, 954–964
- Dunham, I. *et al.* (1999) The DNA sequence of human chromosome 22. *Nature* 402, 489–495
- Mantripragada, K.K. *et al.* (2003) Development of NF2 gene specific, strictly sequence defined diagnostic microarray for deletion detection. *J. Mol. Med.* 81, 443–451
- Schena, M. *et al.* (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270, 467–470
- Pollack, J.R. *et al.* (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. U. S. A.* 99, 12963–12968
- Kauraniemi, P. *et al.* (2001) New amplified and highly expressed genes discovered in the *ERBB2* amplicon in breast cancer by cDNA microarrays. *Cancer Res.* 61, 8235–8240
- Monni, O. *et al.* (2001) Comprehensive copy number and gene expression profiling of the 17q23 amplicon in human breast cancer. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5711–5716
- Porkka, K. *et al.* (2002) Amplification and overexpression of *Elongin C* gene discovered in prostate cancer by cDNA microarrays. *Lab. Invest.* 82, 629–637
- Beheshti, B. *et al.* (2003) Chromosomal localization of DNA amplifications in neuroblastoma tumors using cDNA microarray comparative genomic hybridization. *Neoplasia* 5, 53–62
- Clark, J. *et al.* (2003) Genome-wide screening for complete genetic loss in prostate cancer by comparative hybridization onto cDNA microarrays. *Oncogene* 22, 1247–1252
- Kapranov, P. *et al.* (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296, 916–919
- Rinn, J.L. *et al.* (2003) The transcriptional activity of human Chromosome 22. *Genes Dev.* 17, 529–540
- Mantripragada, K.K. *et al.* DNA copy number analysis of the typically

- deleted region in DiGeorge syndrome using array-CGH with genomic and PCR-based targets. *Int. J. Mol. Med.* (in press)
- 27 Fritz, B. *et al.* (2002) Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res.* 62, 2993–2998
 - 28 Wilhelm, M. *et al.* (2002) Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. *Cancer Res.* 62, 957–960
 - 29 Zhao, J. *et al.* (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
 - 30 Weiss, M.M. *et al.* (2003) Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene* 22, 1872–1879
 - 31 Snijders, A.M. *et al.* (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
 - 32 Sanchez-Izquierdo, D. *et al.* (2003) MALT1 is deregulated by both chromosomal translocation and amplification in B-cell non-Hodgkin lymphoma. *Blood* 101, 4539–4546
 - 33 Martinez-Climent, J.A. *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
 - 34 Locke, D.P. *et al.* (2003) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res.* 13, 347–357
 - 35 Geschwind, D.H. *et al.* (1998) Klinefelter's syndrome as a model of anomalous cerebral laterality: testing gene dosage in the X chromosome pseudoautosomal region using a DNA microarray. *Dev. Genet.* 23, 215–229
 - 36 Yu, W. *et al.* (2003) Development of a comparative genomic hybridization microarray and demonstration of its utility with 25 well-characterized 1p36 deletions. *Hum. Mol. Genet.* 12, 2145–2152
 - 37 Kinzler, K.W. and Vogelstein, B. (1996) Lessons from hereditary colorectal cancer. *Cell* 87, 159–170
 - 38 Bruder, C.E. *et al.* (2001) High resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. *Hum. Mol. Genet.* 10, 271–282
 - 39 Redon, R. *et al.* (2002) Amplicon mapping and transcriptional analysis pinpoint *Cyclin L* as a candidate oncogene in head and neck cancer. *Cancer Res.* 62, 6211–6217
 - 40 Mantripragada, K.K. *et al.* (2003) High-resolution profiling of an 11 Mb segment of human chromosome 22 in sporadic schwannoma using array-CGH. *Int. J. Oncol.* 22, 615–622
 - 41 Wessendorf, S. *et al.* (2003) Hidden gene amplifications in aggressive B-cell non-Hodgkin lymphomas detected by microarray-based comparative genomic hybridization. *Oncogene* 22, 1425–1429
 - 42 Goldstein, M. *et al.* (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
 - 43 Fiegler, H. *et al.* (2003) Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. *J. Med. Genet.* 40, 664–670
 - 44 Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
 - 45 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
 - 46 Armour, J.A. *et al.* (2002) The detection of large deletions or duplications in genomic DNA. *Hum. Mutat.* 20, 325–337
 - 47 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
 - 48 Lopez-Correa, C. *et al.* (2001) Recombination hotspot in *NF1* microdeletion patients. *Hum. Mol. Genet.* 10, 1387–1392
 - 49 Jenne, D.E. *et al.* (2003) Complete physical map and gene content of the human NF1 tumor suppressor region in human and mouse. *Genes Chromosomes Cancer* 37, 111–120
 - 50 Jones, P.A. and Laird, P.W. (1999) Cancer epigenetics comes of age. *Nat. Genet.* 21, 163–167
 - 51 Post, W.S. *et al.* (1999) Methylation of the estrogen receptor gene is associated with aging and atherosclerosis in the cardiovascular system. *Cardiovasc. Res.* 43, 985–991
 - 52 Jirtle, R.L. *et al.* (2000) Genomic imprinting and environmental disease susceptibility. *Environ. Health Perspect.* 108, 271–278
 - 53 Tycko, B. and Ashkenas, J. (2000) Epigenetics and its role in disease. *J. Clin. Invest.* 105, 245–246
 - 54 Robertson, K.D. and Wolffe, A.P. (2000) DNA methylation in health and disease. *Nat. Rev. Genet.* 1, 11–19
 - 55 Laird, P.W. (2003) The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* 3, 253–266
 - 56 Ehrlich, M. *et al.* (2002) Hypomethylation and hypermethylation of DNA in Wilms tumors. *Oncogene* 21, 6694–6702
 - 57 Pollack, J.R. and Iyer, V.R. (2002) Characterizing the physical genome. *Nat. Genet.* 32 (Suppl.), 515–521
 - 58 Zardo, G. *et al.* (2002) Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat. Genet.* 32, 453–458
 - 59 Karpf, A.R. *et al.* (1999) Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14007–14012
 - 60 Jackson-Grusby, L. *et al.* (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* 27, 31–39
 - 61 Liang, G. *et al.* (2002) Analysis of gene induction in human fibroblasts and bladder cancer cells exposed to the methylation inhibitor 5-aza-2'-deoxycytidine. *Cancer Res.* 62, 961–966
 - 62 Suzuki, H. *et al.* (2002) A genomic screen for genes upregulated by demethylation and histone deacetylase inhibition in human colorectal cancer. *Nat. Genet.* 31, 141–149
 - 63 Huang, T.H. *et al.* (1999) Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* 8, 459–470
 - 64 Yan, P.S. *et al.* (2001) Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res.* 61, 8375–8380
 - 65 Yan, P.S. *et al.* (2002) Applications of CpG island microarrays for high-throughput analysis of DNA methylation. *J. Nutr.* 132, 2430S–2434S
 - 66 Tompa, R. *et al.* (2002) Genome-wide profiling of DNA methylation reveals transposon targets of CHROMOMETHYLASE3. *Curr. Biol.* 12, 65–68
 - 67 Gitan, R.S. *et al.* (2002) Methylation-specific oligonucleotide microarray: a new potential for high-throughput methylation analysis. *Genome Res.* 12, 158–164
 - 68 Adorjan, P. *et al.* (2002) Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* 30, e21
 - 69 Ren, B. *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309
 - 70 Iyer, V.R. *et al.* (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409, 533–538
 - 71 Lucito, R. *et al.* (2003) Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* 13, 2291–2305
 - 72 Liu, R.H. *et al.* (2003) Hybridization enhancement using cavitation microstreaming. *Anal. Chem.* 75, 1911–1917