*Chromosome Research* **13**: 237–248, 2005.



# Array CGH technologies and their applications to cancer genomes

Jonathan J. Davies, Ian M. Wilson & Wan L. Lam
*British Columbia Cancer Research Centre, 675 W 10th Ave., Vancouver BC, V5Z 1L3, Canada; Tel: 604-675-8112; Fax: 604-675-8185; E-mail: jdavies@bccrc.ca*

*Key words:* array CGH, cancer genetics, cancer genome, matrix CGH

## Abstract

Cancer is a disease characterized by genomic instability. Comparative genomic hybridization (CGH) is a technique designed for detecting segmental genomic alterations. Recent advances in array-based CGH technology have enabled examination of chromosomal regions in unprecedented detail, revolutionizing our understanding of tumour genomes. A number of array-based technologies have been developed, aiming to improve the resolution of CGH, enabling researchers to refine and define regions in the genome that may be causal to cancer, and facilitating gene discovery at a rapid rate. This article reviews the various array CGH platforms and their use in the study of cancer genomes. In addition, the need for high-resolution analysis is discussed as well as the importance of studying early-stage disease to discover genetic alterations that may be causal to cancer progression and aetiology.

## Introduction

Genetic alteration is a hallmark of cancer. Amplification of oncogenes and deletion of tumour suppressors are common events in cancer progression. Identifying segmental genomic alterations (SeGAs) and the genes they contain will yield molecular targets for diagnostics and therapy. Comparative genomic hybridization (CGH) is a technique designed to identify copy-number changes (Kallioniemi *et al*. 1992). In principle, a normal and a pathological DNA sample are differentially labelled and compared by competitive hybridization against a normal metaphase chromosome spread detecting gains and losses based on changes in signal ratios (Forozan *et al*. 1997). Array-based CGH greatly improves the resolution of the technique by substituting the hybridization target, the metaphase chromosome spread, with genomic segments spotted in an array format (Figure 1). This article reviews the various array CGH platforms and their application to describe cancer genomes.

## Array-based CGH analysis of chromosomal regions altered in cancer

Microarray analysis of genomic copy number changes has led to the development of new, and the adaptation of existing, microarray platforms (Figure 2). Array CGH, also called matrix CGH, was initially used to identify segmental alterations in specific chromosomal regions associated with disease. Solinas-Toldo *et al*. (1997) assessed the use of microarray technology to detect chromosomal imbalances as a way to improve on the resolution of CGH by constructing an array which contained large-insert clones (LIC) spanning 13q14 and other regions. The analysis of a panel of cell line and tumour genomes demonstrated

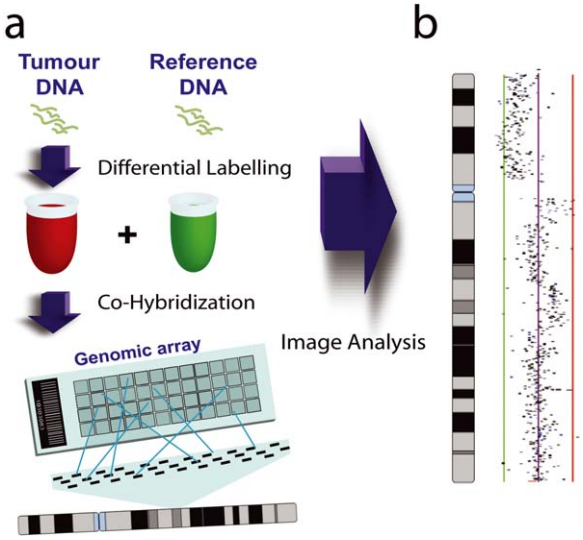



*Figure 1.* General principles of array comparative genomic hybridization. (**a**) Normal and tumour DNA samples are isolated and used to create fluorescently labelled probes, commonly with cyanine 3 (Cy3; green) and cyanine 5 (Cy5; red) dyes. The probes are pooled and competitively cohybridized to a glass slide spotted with a known array of mapped genomic clones. The arrays are analysed with a microarray scanner, producing an image that is used to assess the log2 ratios of the Cy5 to Cy3 intensities for each clone. (**b**) A log2 ratio profile is assembled to determine relative copy number changes between the cancer and tumour samples. Each dot on the graph represents a clone. Values to the left of the '0' line indicate a loss of a genomic region, values to the right indicate a gain or amplification, and values at '0' indicate no change.

that this approach improved the resolution of conventional CGH from 10 Mb to the detection of SeGAs of 75–130 kb in size.

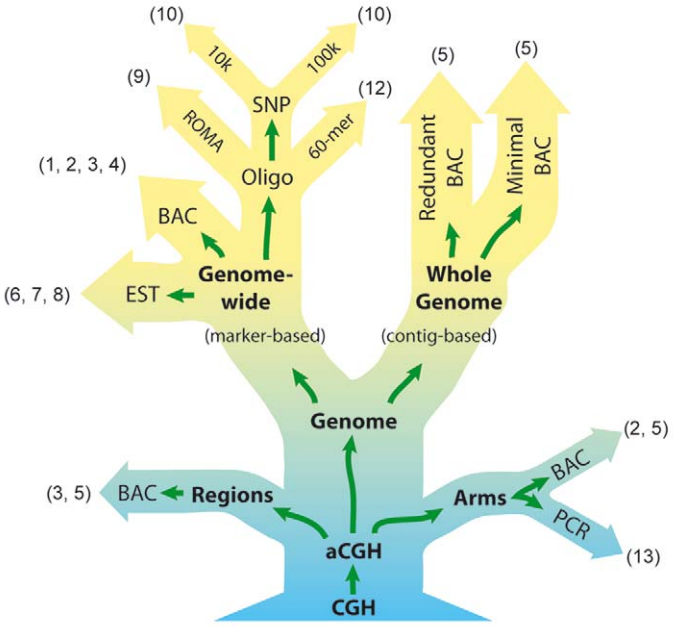
Since this study, a number of regional arrays have been developed to investigate specific genomic 'hotspots' in a variety of diseases. Target clones may span a region often associated with copy number changes in neoplasms or a region harbouring a known gene involved in tumourigenesis in one or more cancers. For example, Garnis and colleagues developed an array covering the ~52-Mb region from 8q21–24 to discover a novel region commonly amplified in pre-invasive oral and lung cancer (Garnis *et al.* 2004c, 2004d, 2004e). Importantly, the amplicon was distinct from the neighbouring *MYC* oncogene and contained a gene whose overexpression appeared to be an early event in cancer progression. Due to the limited resolution of conventional metaphase CGH, previous studies have routinely attributed 8q alterations to the amplification of the *MYC* oncogene. This example illustrates the need for and utility of high-resolution genomic microarrays.

Disease-specific arrays have been developed to survey chromosomal regions commonly altered in chronic lymphocytic leukaemia, mantle cell lymphoma, oral, pancreatic, lung and other cancers, and will be further discussed below (Albertson *et al.* 2000, Massion *et al.* 2002, Garnis *et al.* 2004b, Holzmann *et al.* 2004, Kohlhammer *et al.* 2004, Schwaenen *et al.* 2004).

## Chromosome-specific arrays

The next generation of genomic microarrays examined a specific chromosome or chromosome arm. For example, a chromosome 20 array containing 22 cosmid, P1 phage artificial chromosome (PAC) and bacterial artificial chromosome (BAC) clones as interval markers covering chromosome 20 at 3 Mb resolution (Figure 3; Pinkel. *et al.* 1998). Applying this array to study breast cancer, SeGAs at multiple regions were detected on this chromosome suggesting that higher-density arrays would probably reveal more complex chromosomal alterations in cancer genomes than previously appreciated (Pinkel *et al.* 1998). In another example, Buckley *et al.* (2002) created a comprehensive array for the smallest autosome, chromosome 22. Interestingly, of the 480 target clones spotted, 476 were LICs and four were pools of repeat-free PCR-amplified genomic DNA. The PCR-based probes were used as a proof-of-principle example to show that areas of the genome that are repeat-rich or contain duplications – areas that can give unreliable copy-number data using clone-based assays – can be accurately assessed at a high resolution. This labour-intensive PCR-based method is well suited for regional array construction or to complement higher-coverage arrays on repeat-rich regions (Mantripragada *et al.* 2004). Recently, a number of BAC clone-based contiguous arm arrays have been developed. These arrays provide complete coverage of the 1p, 3p and 5p arms, which are frequently altered in a variety of cancers (Garnis *et al.* 2003, Coe *et al.* 2005, Henderson *et al.* 2005, Garnis *et al.* 2005). These arrays have

a




b

| Probe Type | Manufacturer | Website |
|---|---|---|
| (1) BAC* | Sanger | www.sanger.ac.uk/Projects/Microarrays/ |
| (2) BAC* | UCSF | cc.ucsf.edu/microarray |
| (3) BAC* | DKFZ | www.dkfz-heidelberg.de/kompl_genome |
| (4) BAC | AFCRI | acgh.afcri.upenn.edu |
| (5) BAC | BCCA | www.bccrc.ca/cg/ArrayCGH_Group.html |
| (6) EST | BD Biosciences | www.bdbiosciences.com |
| (7) EST | Bioscience Corp. | www.superarray.com |
| (8) EST | UHN | www.microarray.ca |
| (9) oligo | NimbleGen Systems | www.nimblegen.com/ |
| (10) oligo | Affymetrix | www.affymetrix.com |
| (11) oligo | Bioscience Corp. | www.superarray.com |
| (12) oligo | Agilent Technologies | we.home.agilent.com/USeng/home.html |
| (13) PCR* | UU | puffer.genpat.uu.se/chrom_22_array/chrom22.php |

*Figure 2.* (**a**) 'Evolution' of array comparative genomic hybridization technologies. See text for details. (**b**) Examples of current array platforms. Each platform corresponds to a number in (**a**). *Clone sets contain other large-insert clones. BAC = bacterial bacterial artificial chromosome; EST = expressed sequence tag; UCSF = University of California, San Francisco; DKFZ = Deutsches Krebsforschungszentrum; AFCRI = Abramson Family Cancer Research Institute; BCCRC = British Columbia Cancer Research Centre; UHN = University Health Network; UU = Uppsala Universitet.

proven instrumental in fine-mapping SeGAs in oral squamous cell carcinoma, small-cell and non-small-cell lung carcinoma.

## Genome-wide approaches

Although CGH studies using regional and chromosomal microarrays have yielded a great deal of information, these studies are naturally biased to specific areas of the genome and require *a priori* knowledge of regions of interest. To overcome regional bias, genome-wide arrays were employed (Figure 2). In a pioneer study, Pollack *et al.* (1999) used a cDNA microarray representing 3195 unique cDNA target clones distributed throughout the genome. This study constituted the first array-based genome-wide profiling of human cancer genomes, defining known regions of alteration such as the amplicon containing the

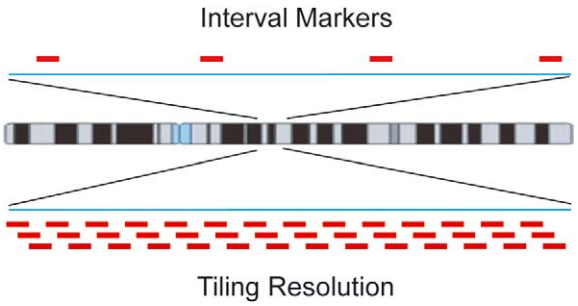



*Figure 3.* Comparison of the genome-target probe densities used to spot microarrays. The resolution of a marker-based microarray is dependent on the distances or gaps between clones – for a hypothetical array with gap sizes of '0', the resolution of the array is limited to the size of the clone. For a tiling-based array, the resolution is dependent on the amount of overlap for each clone, increasing the resolution beyond clone size.

well-characterized *ERBB2* oncogene in breast cancer. The use of cDNA arrays for genome-wide array CGH was a significant advance in the field of cancer genomics. In addition, since the platform was originally designed for gene expression studies, linking genomic data to expression data is greatly facilitated. However, the low signal-to-noise ratio and variable signal intensities are major concerns about using cDNA clones as targets for detecting copy-number alterations. This results from the smaller target size of cDNA clones compared with large-insert genomic clones due to a lack of intronic regions and the varying length of the cDNA target. The technique is effective at resolving high-level SeGAs; however, in order to reliably detect single-copy changes, a moving-average of clone intensities must be calculated, thereby reducing the resolving power of the arrays. Moreover, larger quantities (micrograms) of sample genomic DNA are required in order to generate a robust signal, thus limiting the utility of cDNA array CGH (Pollack et al. 1999). As LICs provide stronger signal intensities, BAC arrays were constructed for genome-wide SeGA profiling. Using 2460 BAC and P1 clones, Snijders *et al.* (2001) created an array with an average marker interval of 1.4 Mb, capable of detecting high- and low-level genomic alterations, substantiating the utility of LIC-based genomic microarrays. Since this initial study, a number of other megabase-interval arrays have been published and are widely used in cancer research (Fiegler *et al.* 2003a, Greshock *et al.* 2004).

Oligonucleotide arrays are also used in copy-number detection (Figure 2). For example, the Affymetrix p501 array, the Mapping 10 K array, and the *Xba*I mapping array 130, containing 8473, 11 555 and 10 043 target probes respectively, have been used in array CGH studies (Bignell *et al.* 2004, Zhao *et al.* 2004). These arrays contain 25-mer oligonucleotides originally designed to assess human single-nucleotide polymorphisms (SNPs). Although such arrays offer an opportunity for integrating LOH data based on SNP analysis, the inherent problem of cross hybridization of oligonucleotide targets to multiple genomic loci invokes the need for complexity reduction of the sample genomic DNA. In order to increase signal-to-noise ratios, the whole-genome sampling assay (WGSA) was developed to greatly reduce the genomic complexity of the sample probes by about 98% to improve hybridization kinetics (Kennedy *et al.* 2003). Briefly, this is achieved by linker-mediated PCR of *Xba*I (or *Eco*RI or *Bgl*II)-digested genomic sample DNA – only short restriction fragments would be amplified (Figure 4).

Hybridization of complexity-reduced probes to SNP arrays is able to detect high-level copy-number changes; however, since the resolution of the method is dependent on the inconsistent

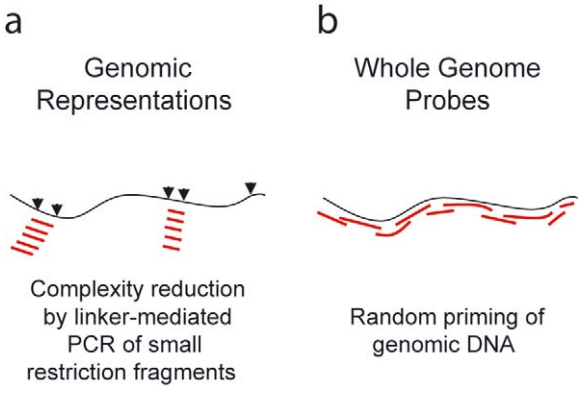



*Figure 4.* Methods used to create genomic probes for array CGH. (**a**) Strategy for generating representational probe by amplification of small restriction fragment in order to reduce genomic complexity of the sample. (**b**) Labelling of total genomic DNA by random priming with degenerate short oligonucleotide primer mixture (e.g. hexamers or octamers).

SNP-density on the array, over a fifth of the chromosomes measured were inconsistent with the SeGAs detected by BAC-array CGH due to poor SNP representation (Bignell *et al.* 2004). In addition, the data generated by the SNP-arrays showed high variability, probably due to the requirement of PCR amplification of genomic sample DNA. This necessitated the use of a moving average to detect copy-number changes, further lowering the effective resolution of the method (Bignell *et al.* 2004). However, as more SNPs are characterized, this method will prove to be a powerful technique in the future (for example, by using a 100k SNP array; Matsuzaki *et al.* 2004). In addition, this method has the advantage of measuring allelic loss of heterozygosity (LOH) alongside copy-number changes using the same platform. Therefore, this technique can greatly improve our understanding of the complex genetic events that may be causal to carcinogenesis.

Similarly, the representative oligonucleotide microarray analysis (ROMA) method reduces the complexity of the genomic DNA sample to ~2.5% of the genome via the *Bgl*II restriction enzyme and linker-mediated PCR amplification (Lucito *et al.* 2003). Using an oligonucleotide microarray consisting of 85 000 target probes of 70-mers designed to detect the genomic 'representations' created by selective amplification of restriction fragments, ROMA aims to detect SeGAs at an average theoretical resolution of 30 kb assuming an even distribution of *Bgl*II sites of the human genome. The identification of submegabase copy-number alterations by this method in breast cancer cell lines and tumours illustrates the need for high-resolution analysis of cancer genomics.

Since ROMA and WGSA employ a genome-reduction step requiring selective PCR amplification of sample DNA, noise is introduced into the assay. Furthermore, the requirement of matched qualities and quantities as well as parallel processing of the sample and test genomes may preclude the use of reduction-based strategies from retrospective studies which rely on archival, formalin-fixed paraffin-embedded material yielding DNA of varying amounts and qualities. In contrast to genomic reduction approaches, Brennan *et al.* (2004) have recently demonstrated that labelling of total genomic DNA can be used to detect single-copy alterations on an oligonucleotide array without a genomic reduction step. However, microgram quantities of starting material are required, which may limit the technique to large tumours and cell lines.

The marker-based genome-wide arrays mentioned thus far, albeit only representing up to 10–15% of the genome, have been instrumental in identifying large (typically greater than 1–2 Mb) variations in somatic genetic changes in tumours as well as in large scale copy number variations in the human population (Iafrate *et al.* 2004, Sebat *et al.* 2004).

## Whole genome tiling path array CGH

In order to comprehensively assess the genome and to identify the focal genetic events occurring during carcinogenesis, a whole genome (as opposed to genome-wide only) approach must be employed (Figure 2). Recently, Ishkanian and colleagues (2004) published the first submegabase resolution tiling-set (*SMRT*) array that contiguously covered the human genome in a tiling path manner (Figure 3). By using overlapping clones, the resolution of the array was increased beyond the size of a single BAC clone and gains and losses of regions as small as 40–80 kb are detectable.

A major advantage of using a tiling-path array is in identifying small (gene level) gains and losses, since marker-based genomic microarrays inherently have a large number of gaps due to the distance between target probes, reducing the likelihood of detecting novel microalterations. That is, the probability of missing a small genetic alteration is inversely proportional to the genome coverage or representation of the detection strategy (Figure 5). The tiling path array offers a much greater probability of detecting small-sized alterations (e.g. 40 kb) than marker-based genomic arrays. The even distribution of markers throughout the genome, as opposed to having a large number of small loci clustered in selected regions, is a key consideration in improving the resolution of genomic scanning strategies. The SMRT array, a tiling-path array, is composed of 32 433 clones spotted in triplicate over two microarray slides (Ishkanian *et al.* 2004). Figure 6 shows a SMRT array CGH karyogram of a lung tumour genome displayed using *SeeGH* software (Chi *et al.* 2004), highlighting the ability of whole genome tiling path arrays in

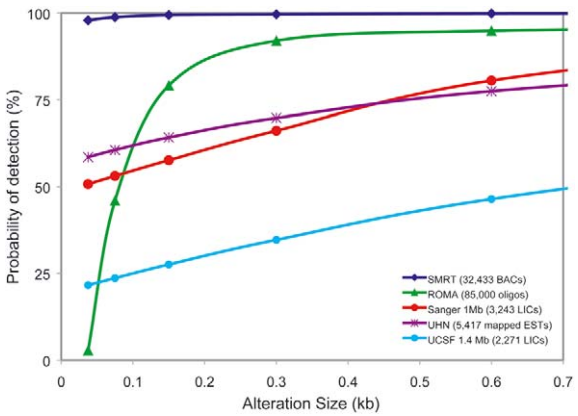



*Figure 5.* The probability of detecting genomic alterations with different array CGH profiling technologies. The probability of a genetic alteration not being detected due to being present in an area not covered by a microarray platform was calculated. For this analysis, we took into account the genomic mapping positions, the sizes of probes for each platform, the locations of telomeres and centromeres, and the criteria each platform uses to determine a region as altered (e.g. the number of clones needed to be altered to consider a region as changed). The number of possible genomic alterations of a given size was calculated for the human genome to determine the percent of possible alterations detectable by each platform. Not surprisingly, the probability of detecting a random alteration decreases for each array type as the size of the alteration decreases. Interestingly, the probability of detecting an alteration of a given size is not only dependent on the number of clones spotted on each array, but also on the size of the clones used to create the array. Larger clones like cDNAs and BACs (average size of 1.5 kb and 150 kb, respectively) have a greater chance of detecting microalterations, while arrays using smaller oligonucleotide probes (<70 bp), as represented by ROMA, have a lower likelihood of detecting alterations below 300 kbp. The SMRT array has the highest probability of detecting small alterations due to the contiguous set of clones used in its construction, with only a small number of interclone gaps which correspond to those in the human genome sequence.

identifying gains and losses as well as the boundaries of the alterations in a single experiment. The clones used have a high degree of overlap and loci redundancy (i.e. multiple clones representing the same regions). This allows for highly reliable detection of copy-number changes with little variation. However, the amount of material required for a two-slide microarray increases the cost of this platform. To reduce the cost of the array synthesis, we have created the SMRT re-array (SMRTr). The SMRTr array contains a more selective set of clones representing 83% of the original collection, eliminating unnecessary redundancy while maintaining tiling path coverage. As a result, the ~27 000 clones can be spotted on to a single slide in duplicate, reducing the cost and time of analysis. These factors are important considerations in establishing whole genome array CGH for use in a clinical setting.

Recently, Bertone *et al.* (2004) constructed a series of 134 tiling arrays using 36-mer oligonucleotide probes representing the non-repetitive fraction of the human genome. This impressive work used ~52 million probes with an average mapped spacing of 46 bp to assess the transcriptional activity throughout the genome, identifying over 10 000 novel expressed sequences. The feasibility of applying this array series to analyse genomic copy-number changes in tumour samples will probably be dependent on the adaptation of DNA amplification techniques due to the amount of sample needed for using the entire array-set.

## Complementary molecular cytogenetic approaches

Array CGH is a method for identifying copy-number alterations but it is not designed for tracking balanced translocations. Spectral karyotyping, also known as multiplex-fluorescence *in-situ* hybridization (FISH), or mFISH, employs chromosome-specific probes to detect gains, losses and rearrangements in a karyogram of colour-coded chromosomes (Garnis *et al.* 2004a). However, microdeletions, inversions and duplications would escape detection by this method (Wang 2002). The resolution limit of spectral karyotyping is reported to be about 3 Mb (Langer *et al.* 2004). A variation of this method uses microdissected chromosomal regions, enabling reverse-FISH mapping against a normal metaphase chromosome spread (Wang 2002, Langer *et al.* 2004). In another variation, chimeric chromosomes isolated by flow sorting are used to paint targets on a genomic microarray (Fiegler *et al.* 2003b). Digital karyotyping is another approach to detect SeGAs. This technique is modelled after serial analysis of gene expression or SAGE. Short nucleotide sequences (tags), adjacent to recognition sites of a restriction enzyme throughout the genome, are enumerated to deduce the relative abundance of tags in a

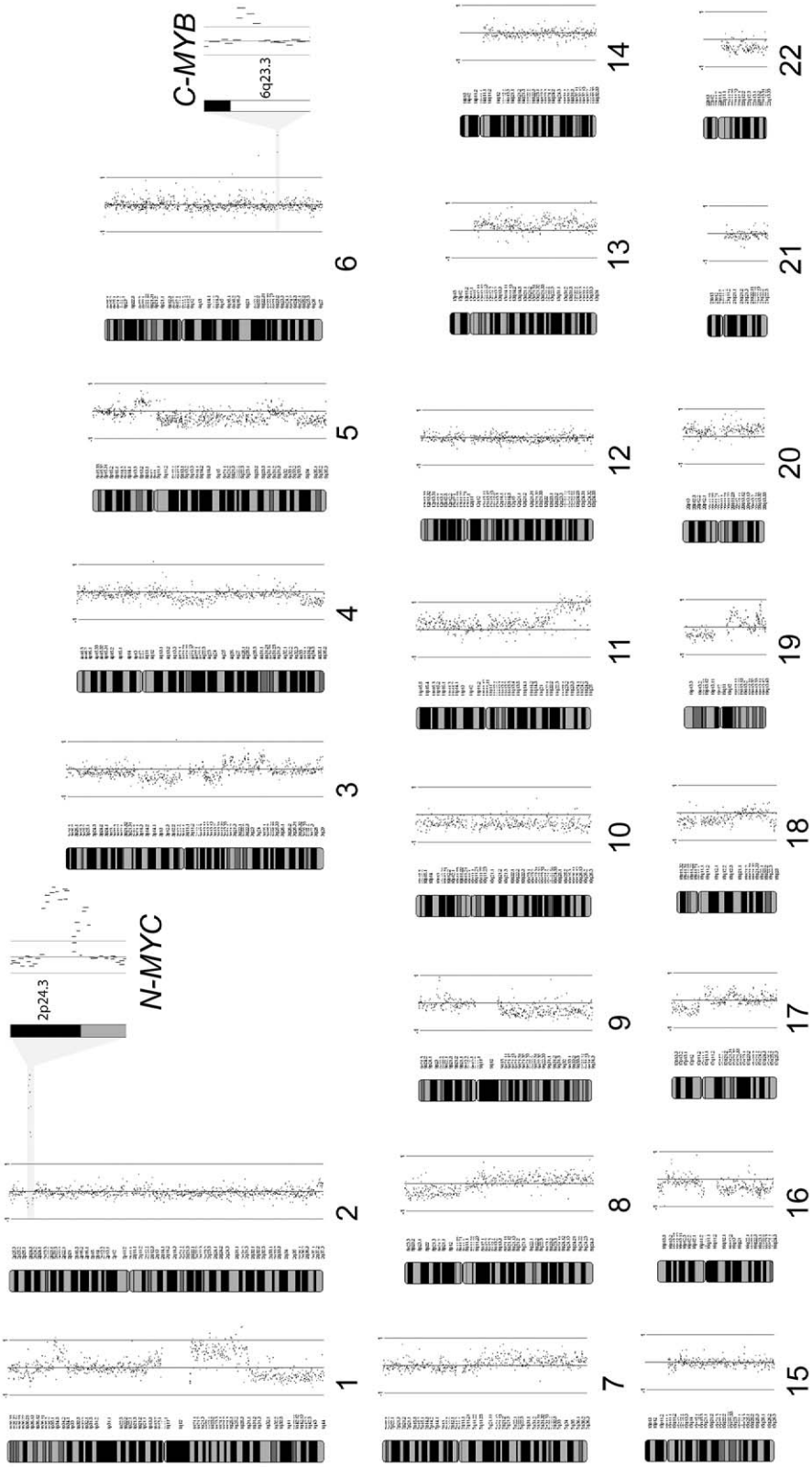



*Figure 6.* Whole-genome profile of a small cell lung cancer sample. Each dot represents a BAC clone and is visualized using *SeeGH* software. Clones around the centre line indicate normal regions in the tumour genome in relation to the normal reference sample. Clones to the left of centre represent homozygous or hemizygous deletions in the tumour, whereas clones to the right represent gains and amplifications. (Note: Array CGH analysis of polyploid tumour genomes against diploid normal samples (e.g. 4n vs. 2n) will appear normal due to image analysis normalization). Focal amplifications of common oncogenes are magnified, highlighting the overlapping nature of the SMRT array.

clone library, reflecting genomic copy number (Wang *et al.* 2002).

## High-resolution view of tumour genomes

To illustrate the need for high-resolution genomic copy number analysis, examples of recent mantle cell lymphoma (MCL) genomic studies are considered. MCL is a B-cell malignancy characterized by t(11;14)(q13;q32). This translocation results in the overexpression of cyclin D1; however, secondary genetic events are necessary for tumourigenesis. Recently, MCL genomes have been well characterized in array-based studies at varying resolution. Using an 812-clone genome-wide marker-based array enriched with regions for MCL classification and 209 clones covering the genome at 15 Mb resolution, Kohlhammer *et al.* (2004) detected an average of 6.7 alterations per case (range 0–16), and detected 50% more SeGAs than identified by conventional CGH analysis of the same samples. Overall, the study identified six regions of recurrent alterations containing candidate genes. A subsequent study performed by Schraders *et al.* (2004) used a 3565-clone genome-wide 800-kb resolution marker-based array with MCL tumour samples. The study detected an average of 15.4 genomic aberrations per sample (range 9–27) and reported 15 recurrently altered regions. Recently, de Leeuw *et al.* (2004) used the whole-genome 32 433-clone tiling-path array to detect a mean of 35.6 alterations per MCL genome (range 21–57), and defined 35 regions of recurrent alteration. Of these recurrent regions, 37% have not been previously described and 26% were less than 1 Mb in size. The above studies of the MCL genome at increasing resolution show a striking, yet anticipated trend in the growing number of SeGAs detected in each sample: arrays with more comprehensive genomic coverage are able to resolve a greater number of alterations. Although the samples used in the de Leeuw *et al.* study were MCL lines that could potentially be subjected to culture artefacts, the detection of 35 alterations per sample remains unprecedented. Furthermore, many of the recurrent regions delineated were submegabase in size, harbouring only 1–2 candidate genes. For example, at tiling resolution, a 2.4-Mb loss at 8p21 initially identified by Kohlhammer *et al.* was further refined to 730 kb containing three candidate tumour suppressor genes including *TNFRSF10B* (tumour necrosis factor receptor superfamily, members 10B). Such regions may go undetected with low-resolution techniques. As with the MCL study, our experience using tiling-set arrays in profiling lung, head & neck, colorectal, breast and prostate tumours suggests that cancer genomes harbour multiple small segmental genomic alterations. These observations reinforce the need to analyse tumour genomes at a resolution far greater than conventional techniques offer.

An important caveat that has gained much attention recently is the unexpected scale of copy-number polymorphisms in the human genome. Studies by Iafrate *et al.* (2004) and Sebat *et al.* (2004) have identified over 200 large-scale copy-number variations throughout the genome, ranging from about 100 kb to >2 Mb in size. As higher-resolution studies are undertaken, the contribution of polymorphisms to the number of 'alterations' detected will need to be addressed. The distinction between inherited polymorphisms and somatic changes can be achieved by careful study design and reference selection, for example, by comparing normal and tumour DNA from the same individuals or by deducing a catalogue of copy number polymorphisms in the human population. However, it may prove worthwhile to determine if a given polymorphism or set of polymorphisms play a role in certain cancer predispositions.

High-resolution analysis of tumour genomes is needed for the discovery of genes involved in the disease. This is evident for two main reasons. First, high-resolution array CGH has the capability to refine known consensus regions of alterations. This is important as it allows researchers to narrow their focus to smaller areas of the genome. For example, trisomy 12 is often found in MCL. In their study, de Leeuw *et al.* (2004) did not see complete trisomy of this chromosome; however, they did discover six distinct recurrent regions of gain over the short and long arm of chromosome 12, one as small as 370 kb containing only two genes. This refinement greatly facilitates and expedites the functional validation of candidate cancer-related genes. Second, high-resolution analysis has a greater probability of detecting small

novel alterations that may be important for the disease but may be missed by lower-resolution techniques (Figure 5). As array-based CGH continues to increase in resolution, cancer genes will continue to become easier to discover. At least in principle, a major hurdle that may limit the feasibility of establishing these techniques for most research laboratories is the inability to manage the vast amounts of data generated. For the same reason, high-resolution arrays may not be practical for clinical use. However, regions discovered using these arrays may be used to create smaller diagnostic arrays that will allow the power of array CGH to be combined with ease of use (discussed below).

## Genetic changes associated with cancer progression

The advancements in array CGH technology have contributed to the study and understanding of underlying genetic changes that correspond to disease progression. Martinez-Climent *et al.* (2003) have used a 2400-clone array with a 1.4 Mb resolution to study follicular centre lymphoma (FCL) cell lines and paired biopsies from patients who have transformed from FCL to the more aggressive diffuse large B-cell lymphoma (DLBCL). The genomic profiles were then compared with their corresponding gene expression data for FCL to DLBCL transformation from another study (Lossos *et al.* 2002). An increase in genomic complexity was observed in association with transformation. Though the expression of a number of genes positively correlated with the type of segmental genomic alteration in which they resided (e.g. overexpression in a gained region), the expression of a larger number of genes within the transformation-identified SeGAs remained unaltered. The heterogeneous mix of SeGAs and the complex correlation between genetic alterations and gene expression lead the authors to conclude that larger data sets will be needed to identify genetic events causal to FCL transformation.

Other studies have focused on using array CGH to describe the genetic events accompanying the progression of premalignant lesions to cancers. Weiss *et al.* (2003) have used a 1.4 Mb interval marker-based array to compare gastric hyperplastic polyps containing intraepithelial neoplasia (dysplasia) to gastric adenomas. The hyperplastic polyps contained numbers of SeGAs per sample comparable to the adenomas. However, both the polyps and adenomas had fewer SeGAs per sample when compared with gastric carcinoma profiles. Interestingly, though the polyps and adenomas contained different SeGAs from each other, the gastric carcinomas contained genetic aberrations common to both premalignant lesion types. These results support the idea that two different origins of gastric cancer exist, generating through either the hyperplastic polyp or the adenoma pathway.

A breast cancer case study by Nyante *et al.* (2004) used a 1.4 Mb interval array to examine the relationship between a lobular carcinoma *in situ* (LCIS) sample, a ductal carcinoma *in situ* (DCIS) sample, and an invasive lobular cancer (ILC) sample from a single patient. Under current breast cancer management protocols, the presence of LCIS is considered to be a risk factor for breast cancer, whereas DCIS is considered a lesion that leads to invasive cancer, resulting in a more aggressive treatment regimen. Yet, the study by Nyante *et al.* (2004) suggests that LCIS and ILC share a genomic signature distinct from the DCIS samples even though LCIS and ILC appear histopathologically different. The presence of this clonality suggests that LCIS can progress to an invasive cancer. However, further studies with a larger sample size are needed to substantiate these claims.

A study using a tiling-path array for the chromosome 1 short arm on preinvasive lung carcinoma *in-situ* samples shows very few recurrent changes on 1p in early lesions (Garnis *et al.* 2005). However, the one recurrent lesion was 0.2 Mb in size and contained the developmental gene, *WNT4.* This led to the investigation of a number of genes involved in the WNT and Notch pathways that were recurrently observed in SeGAs in lung squamous cell carcinomas. Interestingly, a number of these genes were overexpressed in late-stage tumours, supporting the hypothesis that tumours may originate from dysregulated stem cells. Parallel studies on other chromosomes by the same group have demonstrated the value in examining preinvasive lesions as early genetic events may be masked by subsequent changes in invasive cancer (Garnis *et al.* 2004a).

An important and well-known characteristic of most cancer cells is the increase in genetic

instability during disease progression. Indeed, this is highlighted in the studies mentioned: all later-stage tumours had a greater average number of SeGAs per sample than their precursors. Though many of these secondary genetic changes may be responsible for disease maintenance and proliferation, the majority are not likely to contribute to pathogenesis and are probably a consequence of random events. This is suggested in the study by Martinez-Climent *et al*. (2003), in which the majority of the genes within the identified SeGAs had unaltered expression levels. Discovering frequently altered regions may reveal important disease-related genes; however, these genes are not necessarily causal to tumour formation. Studying pre-malignant or precursor lesions helps to avoid these issues. The increased genomic stability of tumour precursors improves the likelihood that a SeGA discovered may be necessary for the disease. Furthermore, discovering localized regions that contain these genes might be easier, as they may not be 'masked' by the gross genetic alterations occurring at later stages. Clearly, the data generated by these precursor studies is valuable. This is even more evident when one considers the rarity of these lesions. For example, a practicing clinician may only encounter a lung carcinoma *in situ* once every few years (Stephen Lam, personal communication). As detection techniques improve and screening programmes are implemented, the chance to study these precious lesions will increase, which will assuredly generate a better understanding of cancer progression and aetiology.

## Clinical translation

At this point, array CGH technology has been used primarily as a tool for discovering genetic alterations. Relating alterations to clinical features could yield markers predictive of disease behaviour. For example, the correlation of mantle cell lymphoma matrix CGH data with clinical course showed that 8p and 13q14 deletions conferred poor prognosis (Kohlhammer *et al.* 2004). In a recent report, Lichter and colleagues developed a chronic lymphocytic leukaemia array to demonstrate the potential in adapting this technology as an automated tool in clinical classification (Schwaenen *et al*. 2004).

The correlation of cytogenetic aberrations with disease outcome has already proven to be reliable and utilized in guiding treatment. Somatic mutations are stable unlike gene expression changes which may be transient. Detection of changes in tumour DNA, such as loss of heterozygosity, gene amplification and microsatellite instability, became useful markers associated with malignant development. Historically, technological developments such as fluorescence *in-situ* hybridization and comparative genomic hybridization greatly increased the resolution power of the cytogenetic approach with concomitant benefits. Significantly, array CGH has the potential to improve this resolution by orders of magnitude at a reasonable cost and offers the capacity to be high throughput. This is especially applicable to the analysis of solid tumours where *in-vitro* cell culture for karyotype analysis is difficult. The major challenges in adapting this technology for a clinical setting are making the technology: (1) robust and simple to use which would require development in instrumentation and analytical software, and (2) price competitive with the current clinical cytogenetic techniques.

## Acknowledgements

The authors would like to thank Bradley Coe for critical discussion and Cathie Garnis, William Lockwood, and Timon Buys for proofreading. We wish to apologize for our inability to include all cancer array CGH studies in this review due to space constraints. For a larger list of recommended reading, please see http://www.bccrc.ca/cg/ArrayCGH_Group.html. This work was supported by funds from Genome Canada/British Columbia, National Institute of Dental and Cranial Research RO1 DE 015965-01, Lymphoma Research Foundation, Canadian Institute of Health Research, and Natural Sciences and Engineering Research Council of Canada scholarship to J.J.D.

## References

Albertson DG, Ylstra B, Segraves R *et al*. (2000) Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat Genet* **25**: 144–146.

Bertone P, Stolc V, Royce T *et al*. (2004) Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**: 2242–2246.

Bignell GR, Huang J, Greshock J *et al*. (2004) High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* **14**: 287–295.

Brennan C, Zhang Y, Leo C *et al*. (2004) High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res* **64**: 4744–4748.

Buckley PG, Mantripragada KK, Benetkiewicz M *et al*. (2002) A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. *Hum Mol Genet* **11**: 3221–3229.

Chi B, DeLeeuw RJ, Coe BP, MacAulay C, Lam WL (2004) SeeGH—a software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics* **5**: 13, 1–6.

Coe BP, Henderson LJ, Garnis C *et al*. (2005) High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines. *Genes Chromosomes Cancer* **42**: 308–318.

de Leeuw RJ, Davies JJ, Rosenwald A *et al*. (2004) Comprehensive whole genome array CGH profiling of mantle cell lymphoma model genomes. *Hum Mol Genet* **13**: 1827–1837.

Fiegler H, Carr P, Douglas EJ *et al*. (2003a) DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* **36**: 361–374.

Fiegler H, Gribble SM, Burford DC *et al*. (2003b) Array painting: a method for the rapid analysis of aberrant chromosomes using DNA microarrays. *J Med Genet* **40**: 664–670.

Forozan F, Karhu R, Kononen J, Kallioniemi A, Kallioniemi OP (1997) Genome screening by comparative genomic hybridization. *Trends Genet* **13**: 405–409.

Garnis C, Baldwin C, Zhang L, Rosin MP, Lam WL (2003) Use of complete coverage array comparative genomic hybridization to define copy number alterations on chromosome 3p in oral squamous cell carcinomas. *Cancer Res* **63**: 8582–8585.

Garnis C, Buys TP, Lam WL (2004a) Genetic alteration and gene expression modulation during cancer progression. *Mol Cancer* **3:** 9, 1–23.

Garnis C, Campbell J, Zhang L, Rosin MP, Lam WL (2004b) OCGR array: an oral cancer genomic regional array for comparative genomic hybridization analysis. *Oral Oncol* **40**: 511–519.

Garnis C, Coe BP, Ishkanian A, Zhang L, Rosin MP, Lam WL (2004c) Novel regions of amplification on 8q distinct from the MYC locus and frequently altered in oral dysplasia and cancer. *Genes Chromosomes Cancer* **39**: 93–98.

Garnis C, Coe BP, Zhang L, Rosin MP, Lam WL (2004d) Overexpression of LRP12, a gene contained within an 8q22 amplicon identified by high-resolution array CGH analysis of oral squamous cell carcinomas. *Oncogene* **23**: 2582–2586.

Garnis C, MacAulay C, Lam S, Lam W (2004e) Genetic alteration on 8q distinct from MYC in bronchial carcinoma in situ lesions. *Lung Cancer* **44**: 403–404.

Garnis C, Campbell J, Davies JJ, Macaulay C, Lam S, Lam WL (2005) Involvement of multiple developmental genes on chromosome 1p in lung tumourigenesis. *Hum Mol Genet* **14**: 475–482.

Greshock J, Naylor TL, Margolin A *et al*. (2004) 1-Mb resolution array-based comparative genomic hybridization using a BAC clone set optimized for cancer gene analysis. *Genome Res* **14**: 179–187.

Henderson LJ, Coe B P, Lee E *et al*. (2005) Genomic and gene expression profiling of minute alterations of chromosome arm 1p in small cell lung carcinoma cells. *Br J Cancer* (in press).

Holzmann K, Kohlhammer H, Schwaenen C *et al*. (2004) Genomic DNA-chip hybridization reveals a higher incidence of genomic amplifications in pancreatic cancer than conventional comparative genomic hybridization and leads to the identification of novel candidate genes. *Cancer Res* **64**: 4428–4433.

Iafrate AJ, Feuk L, Rivera MN *et al*. (2004) Detection of large-scale variation in the human genome. *Nat Genet* **36**: 949–951.

Ishkanian AS, Malloff CA, Watson SK *et al*. (2004) A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* **36**: 299–303.

Kallioniemi A, Kallioniemi OP, Sudar D *et al*. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumours. *Science* **258**: 818–821.

Kennedy GC, Matsuzaki H, Dong S *et al*. (2003) Large-scale genotyping of complex DNA. *Nat Biotechnol* **21**: 1233–1237.

Kohlhammer H, Schwaenen C, Wessendorf S *et al*. (2004) Genomic DNA-chip hybridization in t(11;14)-positive mantle cell lymphomas shows a high frequency of aberrations and allows a refined characterization of consensus regions. *Blood* **104**: 795–801.

Langer S, Kraus J, Jentsch I, Speicher MR (2004) Multicolor chromosome painting in diagnostic and research applications. *Chromosome Res* **12**: 15–23.

Lossos IS, Alizadeh AA, Eisen MB *et al*. (2002) Ongoing immunoglobulin somatic mutation in germinal center B cell-like but not in activated B cell-like diffuse large cell lymphomas. *Proc Natl Acad Sci USA* **99**: 8886–8891.

Lucito R, Healy J, Alexander J *et al*. (2003) Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* **13**: 2291–2305.

Mantripragada KK, Tapia-Paez I, Blennow E, Nilsson P, Wedell A, Dumanski JP (2004) DNA copy-number analysis of the 22q11 deletion-syndrome region using array-CGH with genomic and PCR-based targets. *Int J Mol Med* **13**: 273–239.

Martinez-Climent JA, Alizadeh AA, Segraves R *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.

Massion PP, Kuo WL, Stokoe D *et al*. (2002) Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: implications of the phosphatidylinositol 3-kinase pathway. *Cancer Res* **62**: 3636–3640.

Matsuzaki H, Dong S, Loi H *et al.* (2004) Genotyping over 100 000 SNPs on a pair of oligonucleotide arrays. *Nature Meth* **1**: 109–111.

Nyante SJ, Devries S, Chen YY, Hwang ES (2004) Array-based comparative genomic hybridization of ductal carcinoma *in situ* and synchronous invasive lobular cancer. *Hum Pathol* **35**: 759–763.

Pinkel D, Segraves R, Sudar D *et al.* (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**: 207–211.

Pollack JR, Perou CM, Alizadeh AA *et al.* (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**: 41–46.

Schraders M, Pfundt R, Straatman HM *et al.* (2004) Novel chromosomal imbalances in mantle cell lymphoma detected by genome-wide array-based comparative genomic hybridization. *Blood* **105**: 1686–1693.

Schwaenen C, Nessling M, Wessendorf S *et al.* (2004) Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. *Proc Natl Acad Sci USA* **101**: 1039–1044.

Sebat J, Lakshmi B, Troge J *et al.* (2004) Large-scale copy number polymorphism in the human genome. *Science* **305**: 525–528.

Snijders AM, Nowak N, Segraves R *et al.* (2001) Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* **29**: 263–264.

Solinas-Toldo S, Lampel S, Stilgenbauer S *et al.* (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**: 399–407.

Wang N (2002) Methodologies in cancer cytogenetics and molecular cytogenetics. *Am J Med Genet* **115**: 118–124.

Wang TL, Maierhofer C, Speicher MR *et al.* (2002) Digital karyotyping. *Proc Natl Acad Sci USA* **99**: 16156–16161.

Weiss MM, Kuipers EJ, Postma C *et al.* (2003) Genome wide array comparative genomic hybridisation analysis of premalignant lesions of the stomach. *Mol Pathol* **56**: 293–298.

Zhao X, Li C, Paez JG *et al.* (2004) An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* **64**: 3060–3071.