

REVIEW

# Recent advances in array comparative genomic hybridization technologies and their applications in human genetics

William W Lockwood<sup>\*,1,2</sup>, Raj Chari<sup>1,2</sup>, Bryan Chi<sup>1</sup> and Wan L Lam<sup>1</sup>

<sup>1</sup>*Cancer Genetics and Developmental Biology, British Columbia Cancer Research Centre, Vancouver BC, Canada V5Z 1L3*

**Array comparative genomic hybridization (array CGH) is a method used to detect segmental DNA copy number alterations. Recently, advances in this technology have enabled high-resolution examination for identifying genetic alterations and copy number variations on a genome-wide scale. This review describes the current genomic array platforms and CGH methodologies, highlights their applications for studying cancer genetics, constitutional disease and human variation, and discusses visualization and analytical software programs for computational interpretation of array CGH data.**

*European Journal of Human Genetics* (2006) **14**, 139–148. doi:10.1038/sj.ejhg.5201531; published online 16 November 2005

**Keywords:** array CGH; matrix CGH; cancer genome; cancer genetics; genetic disease; copy number alteration; array CGH software

## Introduction

Comparative genomic hybridization (CGH) is a method designed for identifying chromosomal segments with copy number aberration. Differentially labeled genomic DNA samples are competitively hybridized to chromosomal targets, where copy number balance between the two samples is reflected by their signal intensity ratio. Since its development in the early 1990s, a great deal of effort has been devoted to improving the resolution of this technology. The use of DNA targets immobilized in an array format, replacing the conventional metaphase chromosome spreads, represents a significant advance.<sup>1,2</sup> Traditionally, the resolution of array-based CGH has been defined by the genomic distance between each DNA target represented on the array.<sup>3,4</sup> Pollack *et al*<sup>5</sup> extended this technology to facilitate high-resolution genome-wide

survey of segmental alterations by using cDNA microarrays for CGH analysis. The development of a whole-genome bacterial artificial chromosome (BAC) tiling path array has further improved the resolution of CGH.<sup>6</sup> This article describes the various types of genomic arrays, highlights their application in identifying genetic alterations in cancer and genetic diseases, and summarizes computational software used in the visualization and analysis of array CGH data.

## Genomic microarrays

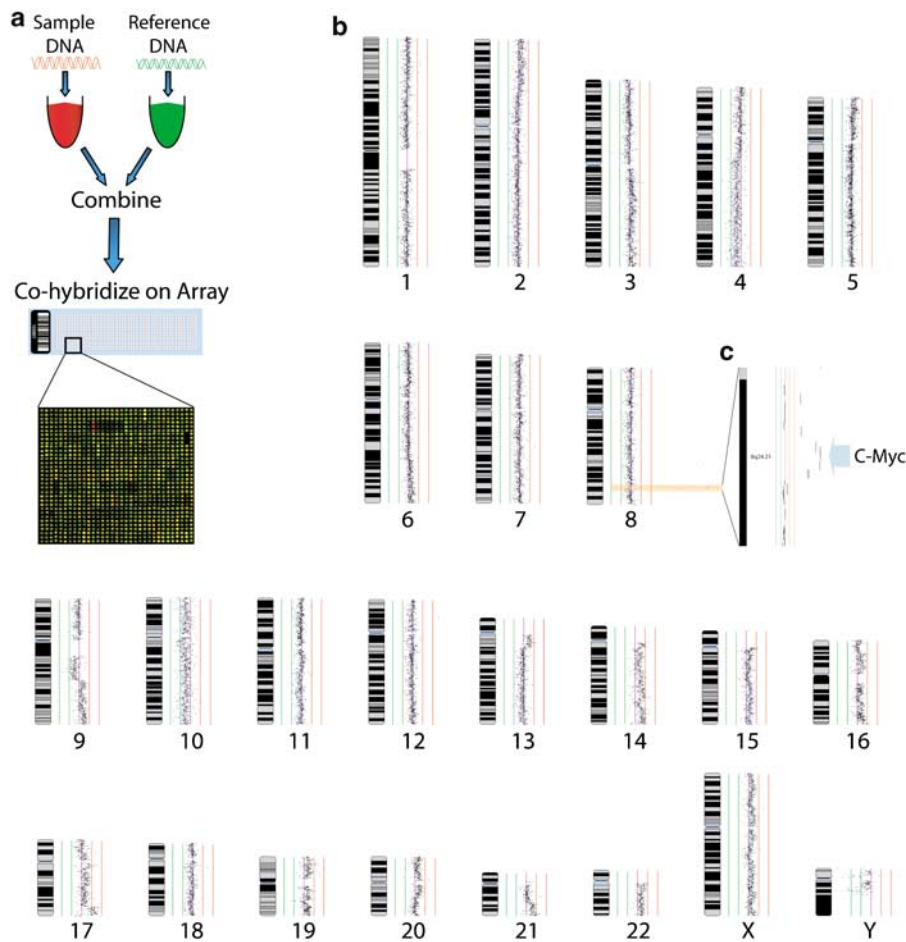
Although numerous platforms have been developed to support array CGH studies, they all revolve around a common principle of detecting copy number alterations between two samples (Figure 1). These platforms vary in terms of the size of the genomic elements spotted and their coverage of the genome. This section characterizes commonly used genome-wide approaches (Supplemental Table 1) and highlights their relevant features.

## Genome-wide marker-based arrays

The genome-wide approach to array CGH was introduced using cDNA microarrays, which were originally used in

\*Correspondence: W Lockwood, BC Cancer Research Centre, 675 West 10th Avenue, Vancouver, BC, V5Z 1L3, Canada.  
Tel: +1 604 675 8112; Fax: +1 604 675 8185;  
E-mail: wlockwood@bccrc.ca

<sup>2</sup>These two authors contributed equally to this work  
Received 8 June 2005; revised 20 September 2005; accepted 13 October 2005; published online 16 November 2005



**Figure 1** Principles of array comparative genomic hybridization. (a) Sample and reference DNA are differentially labeled with fluorescent dyes (typically cyanine-3 and cyanine-5), combined, and cohybridized to a microarray containing spots of genomic material. The sample and reference competitively bind to the spots and the resulting fluorescence intensity ratios are reflected by their relative quantities. (b) Whole-genome ideogram of a small cell lung cancer cell line hybridized against a normal male reference on the submegabase resolution tiling array. Each black dot represents a single BAC clone spotted on the array. The red, purple and green vertical lines adjacent to each chromosome represent log<sub>2</sub> fluorescence ratios of 0.5, 0 and -0.5, respectively. (c) Magnified view of a high-level amplification at the *c-Myc* oncogene locus at 8q24.21 in the small cell lung cancer cell line.

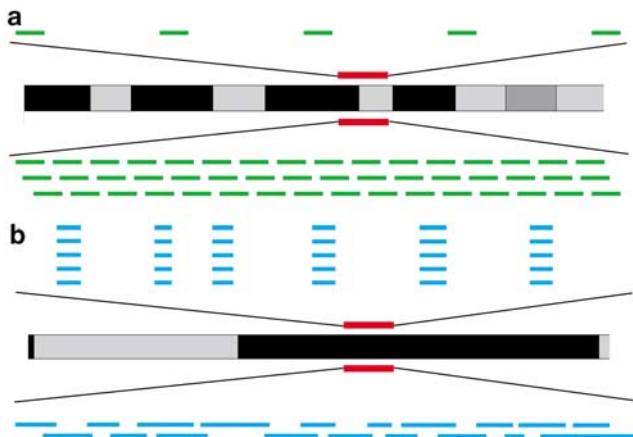
gene expression profiling.<sup>5</sup> One advantage of this technique is that high-level amplifications and deletions can be directly correlated to expression changes using the same platform.<sup>5,7</sup> However, only exonic regions of the genome are covered by the cDNA targets making alterations to promoter regions and other protein binding sites undetectable. New generation cDNA arrays are consisted of exon-specific targets.<sup>8</sup>

Marker-based large insert clones (LIC) arrays sample the genome at megabase intervals, typically covering about 10% of the genome<sup>9–12</sup> (Figure 2a). However, these arrays are often labeled as ‘high resolution’, relative to classical chromosomal CGH analysis. The main advantages of genome-wide arrays are that LICs, such as BACs, provide robust targets for sensitive detection of hybridization signals and that BACs are not limited to loci annotated with genes. The size of the arrayed elements also provide a

higher signal to noise ratio compared to platforms using smaller targets as signal intensities increase as the complexity of the DNA spotted increases.<sup>13</sup> Thus, BAC-based platforms allow highly sensitive and reproducible detection of a wide range of copy number changes including single copy number gains and losses, homozygous deletions and high-level amplifications.<sup>13</sup>

#### Use of single nucleotide polymorphisms oligonucleotide arrays in CGH

Arrays of photolithographically synthesized short oligonucleotides (21–25 nucleotides in length) originally designed for detecting single nucleotide polymorphisms (SNP), have been recruited for use in copy number assessment in CGH experiments.<sup>14</sup> In a method known as whole-genome sampling assay (WGSA), linker-mediated PCR is performed on the sample DNA to enrich for small *Xba*I restriction



**Figure 2** Genomic array and sample labeling design (a) display of marker based (top) and tiling path (bottom) approaches to array design. Marker-based approaches sample the genome at intervals, while the tiling path approach improves resolution by using overlapping clones. (b) Illustrates genomic representation (top) and whole-genome (bottom) approaches to sample labeling. The representational approach enriches for short restriction fragments by linker-mediated PCR amplification while whole-genomic labeling typically involves random priming of the genomic DNA sample without complexity reduction.

fragments throughout the genome in order to reduce sample complexity prior to hybridization<sup>15</sup> (Figure 2b). Although the reduced sample no longer represents the entire genome, this process decreases the probability of crosshybridization to multiple short oligonucleotide targets on the array, effectively decreasing nonspecific signals.<sup>14</sup> The strength of this strategy is its ability to relate copy number and allelic status at selected loci.

Using cancer cell lines, Bignell *et al*<sup>14</sup> compared the performance of WGS coupled with an array consisting of 8473 predicted SNPs with that of conventional BAC array CGH hybridized with the same samples. Although high-level amplifications and homozygous deletions were evident, the oligonucleotide array showed greater variation in the detection of single copy gains and losses in contrast to the BAC array.<sup>14</sup> In another study, Zhao *et al*<sup>16</sup> compared SNP, cDNA and BAC arrays for their ability to detect copy number changes in the breast cancer cell line BT474. These array platforms detected a similar but not identical pattern of alterations across the genome, with SNP results showing 70% similarity to BAC arrays and 62% similarity to the cDNA method.<sup>16</sup> The BAC arrays showed the highest signal to noise ratios, making them better suited to detect single copy alterations.<sup>16</sup> However, the SNP arrays allow copy number changes and genotype to be measured in a single experiment. The recent development of high-density SNP arrays, for example the Affymetrix GeneChip Mapping 100K array, will improve copy number assessment at the SNP loci residing within the genome-wide linker-mediated PCR amplified restriction fragments.<sup>17</sup>

### Long oligonucleotide arrays

Increasing the length of the target oligonucleotide aims to improve hybridization specificity. Unlike SNP array platforms, arrays of spotted oligonucleotides of typically 60–70mers in length are used to directly assay genomic DNA samples without the need for a complexity reduction step prior to hybridization.<sup>18–20</sup> The ratios detected on a genome-wide scale have been reported to be comparable to BAC arrays in the magnitude of signal and background noise.<sup>18</sup> Recently, a tiling path oligonucleotide array with 6kb median probe spacing was utilized to analyze chromosomal breakpoints in neuroblastoma.<sup>21</sup> However, these array platforms typically require the calculation of a moving average to observe single copy changes which may decrease their effective resolution.<sup>19</sup> Future studies are needed to determine the effectiveness of these techniques for use with archival clinical samples.

Sample genome complexity reduction was combined with hybridization to long oligonucleotide arrays in a method called representational oligonucleotide microarray analysis (ROMA). A genomic DNA sample is cleaved with a methylation insensitive restriction enzyme (usually *Bgl*II) followed by linker-mediated PCR which enriches for fragments <1.2kb in length.<sup>22</sup> This results in a low complexity representation comprising ~2.5% of the genome, which improves the signal to noise ratio when hybridized to oligonucleotide targets. The arrays are comprised of 70mer probes designed to hybridize to a specific representation fragment from the genome.<sup>22</sup> These oligonucleotides are from random portions of the genome and are picked according to their signal strength. This leads to variable coverage across the genome with some areas poorly represented and others densely represented. Currently, arrays with approximately 85000 probes have been developed allowing amplifications and single copy deletions in cancer cell line genomes to be observed.<sup>22</sup>

### Whole-genome tiling path array

Although the methods described above allow copy number changes to be assessed on a genome-wide scale, the coverage of the arrayed elements can vary greatly. This leads to large gaps where no genomic information is obtainable. Thus, to fully understand the alterations occurring in various diseases, probes covering the entire genome are required. To date, Ishkanian *et al*<sup>6</sup> have produced the only array CGH platform with whole-genome coverage (Figure 1). This submegabase resolution tiling set (SMRT) array is comprised of 32433 overlapping BAC clones spotted in triplicate on two glass slides. Like other large insert clone-based approaches, the SMRT array yields high signal to noise ratios due to the hybridization sensitivity of the BACs to their corresponding genome targets. In contrast to marker-based approaches, the overlapping arrangement of the BAC clones abrogates the need to infer genetic events between marker clones and the

redundancy provides confirmation of copy number status at each locus (Figure 2). The tiling nature also increases the probability of detecting microalterations that may fall between marker probes in other array platforms.

However, the major consideration in interpreting whole-genome BAC array data is the fact that some of the clones map to multiple places in the genome due to crosshybridization to highly homologous sequences. In megabase interval arrays, these clones would be excluded; however, in order to provide tiling path coverage, such clones have to be included. Tracking of these clones computationally would improve the accuracy of interpretation.

### Sequencing-based alternatives for copy number analysis

Digital karyotyping and fosmid paired-end sequencing are emerging methods for genome-wide profiling of copy number variations.<sup>23,24</sup> Digital karyotyping involves the isolation and enumeration of short sequence tags corresponding to specific loci, and tag abundance reflects copy number status.<sup>23</sup> Similarly, fosmid paired-end sequencing enumerates the relative abundance of cloned fragments created from a genome of interest. However, by aligning the end sequences against the human genome sequence assembly, one can also detect structural variations such as insertions, deletions and inversions.<sup>24</sup> Despite the precision of these sequence-based technologies, their widespread use will likely require a reduction in costs associated with DNA sequencing.

### Array requirements and considerations

The choice of platform technology for an array CGH study primarily depends on the type of samples being analyzed and the level of detail desired. Here, we examine issues pertaining to input material as well as genomic resolution.

### Quality and quantity of sample DNA

A major consideration in selecting an array platform is sample requirement. DNA quantity may be limiting when analyzing small biopsies, while DNA quality may be compromised in formalin-fixed, paraffin-embedded archival specimens. Large insert clone (such as those comprised of BAC clones of ~150 kb in size) arrays efficiently capture signals from samples of low DNA quantity and quality for genome-wide analysis, while oligonucleotide and small PCR fragments could facilitate more detailed investigation at selected regions when DNA quality and quantity are not limiting. BAC arrays require 200–400 ng of DNA,<sup>2,6</sup> whereas oligonucleotide and cDNA platforms typically require microgram amounts.<sup>5,18,25</sup> Amplification techniques, such as those used in WGS and ROMA, have proven effective in increasing hybridization signal strength and limiting noise through the reduction of sample complexity at the expense of genomic coverage. However, these

techniques yield variable results in repeatability using the same sample and the biases potentially introduced by the PCR step are not fully understood.<sup>18</sup>

Tissue heterogeneity in a sample affects detection sensitivity of copy number changes and therefore is another consideration for array selection.<sup>26</sup> Noncancerous cells in a tumor sample effectively dampen the shift in signal ratio associated with genetic alterations in the cancer cells. Garnis *et al*<sup>27</sup> mimicked this phenomenon experimentally and concluded that increasing the number of measurements over a genomic distance could provide more data points within a segmental alteration, thereby increasing the probability of detection. The use of tiling path resolution arrays, as opposed to interval marker arrays, should be considered in analyzing heterogeneous tissue samples.

Another important consideration in array CGH experiments is the selection of reference DNA. Common options are using a sex-matched reference (eg male reference for male sample), sex-mismatched, a reference consisting of DNA from a pool of individuals or using a reference from a single individual.

### Functional resolution of genome-wide array CGH

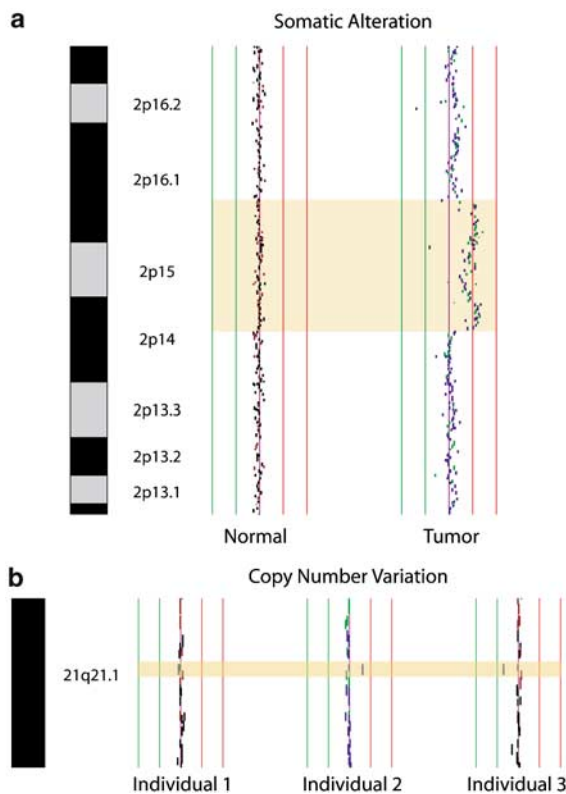
The definition of resolution in terms of array CGH is ambiguous. A practical definition is the genomic distance between array elements (clones or oligonucleotides). However, such elements may not be evenly distributed throughout the genome and some platforms may require multiple elements to detect an alteration, so that calculating resolution based on the mean or median distance would be misleading. Furthermore, tiling path arrays that span chromosomes with overlapping clones cannot be assessed in this way. A functional measure of resolution can be the size limit of detecting a segmental copy number alteration.<sup>28</sup>

### Applications of array CGH

Although the most frequent use of array CGH is in the detection of somatic segmental changes pertaining to cancer (Figure 3a), there are other applications. We now review the use of array CGH in measuring copy number status in cancer, in genetic diseases, and in evolutionary comparisons. We also consider its potential as a diagnostic tool in a clinical setting.

### Identifying somatic DNA alterations in cancer

In the past 5 years, there have been numerous reports of high-resolution array CGH studies of copy number alterations initially focusing on specific regions of tumor genomes and later expanding to entire chromosome arms, for example, a 3p array of a tiling set of 535 BAC clones used in defining common gains and losses in oral cancer;<sup>29</sup> a 5p array of 491 BAC clones spanning the 50 Mb of 5p<sup>28</sup> and a 1p array of 642 ordered BACs spanning 120 Mb<sup>30</sup> that



**Figure 3** Somatic alterations and copy number variations. (a) Example of a segmental duplication observed at chromosome arm 2p present in the cancer cells but absent in the normal cells from the same individual. Each black dot represents a single BAC clone spotted on the array. The purple line represents equal fluorescent intensity ratio between sample and reference. Copy number gain (and loss) shifts the ratio to the right (and left). (b) Illustrates a copy number variation observed at chromosomal region 21q21.1. Three normal individuals exhibit equal, more and fewer copies relative to the reference DNA, indicating variation in the population.

were instrumental in the discovery of genes involved in lung tumorigenesis.<sup>28,30–32</sup> Furthermore, arm-specific arrays have been used to profile astrocytic tumors and other various cancers.<sup>33,34</sup>

In terms of genome-wide approaches, cDNA microarrays and interval LIC arrays have yielded much information on the genomic landscape of a variety of cancers and the discovery of recurrent genetic alterations.<sup>5,35–38</sup> In addition, genome-wide profiles are used to deduce features for disease classification including drug response.<sup>39–43</sup>

The recent development of a whole-genome tiling path array has advanced such analysis to examining tumor genomes at unprecedented details.<sup>6</sup> Whole-genome profiling of mantle cell lymphoma (MCL) cell lines using the SMRT array has identified an average of 35 alterations per genome with equal numbers of gains and losses and found recurrent alterations as small as 130 kb in size.<sup>44</sup> Further utilization of the SMRT array has provided useful in mapping focal amplifications in oral and lung cancer as

well as osteosarcoma.<sup>38,45–47</sup> Application of array CGH to epithelial and hematological malignancies have yielded novel genetic alterations that have escaped the detection of conventional methods, and facilitated a concerted search for multiple disruptions in biological pathways.<sup>42,43</sup>

### Identifying segmental copy number changes in genetic diseases

Segmental duplications and deletions have been well documented in inherited diseases.<sup>10,48,49</sup> Advances in array-based CGH have greatly facilitated the discovery of such genetic alterations. Megabase interval genomic arrays have been instrumental in delineating regions affected in a variety of genetic diseases. Submicroscopic chromosomal deletions and duplications were identified in cytogenetically normal patients exhibiting mental retardation and dysmorphisms.<sup>10,48</sup> In addition, copy number changes were refined in Cri-du-chat syndrome, congenital diaphragmatic hernia (CDH), and Prader-Willi Syndrome (PWS) using array CGH.<sup>50–52</sup> These examples illustrate the value of array CGH in identifying aberrations that have escaped traditional cytogenetic analysis and in refining aberrations previously characterized in various disorders.

### Identifying DNA copy number variation in the human population

A new area in which array CGH is being utilized is in the characterization of large-scale DNA variations (Figure 3b). Using an array of 2632 large-insert clones, 55 unrelated individuals were examined to quantify genetic variation.<sup>53</sup> Overall, 255 loci across the human genome contained genomic imbalances, 24 of them were present in over 10% of the participants and six of these large-scale copy number variations (LCVs) were present in 20% of the individuals. Among the total number of LCVs, over half (142) harbored genes. Strikingly, 14 LCVs were located near loci associated with cancer or genetic diseases, suggesting that certain individuals may have higher susceptibility to disease than others. A subsequent study using a BAC array specific to potential rearrangement hotspots in the genome was used to assess copy number variation in 41 normal individuals.<sup>54</sup>

Similar conclusions were derived using the ROMA technology. In total, 20 subjects were analyzed and 76 unique copy number variations were discovered.<sup>55</sup> From a total of 226 copy number differences, on average, each individual harbored 11 variations with a length of 465 kb. The copy number variations contain genes that have been implicated in cell growth and other functions.

Clearly, these three studies illustrate the utility of array CGH in investigating large-scale variation. However, due to the limited genomic coverage of the techniques used, more comprehensive studies using whole-genome tiling path arrays are necessary to enumerate and identify all such LCVs in the human population.

### Evolutionary characterization

Array CGH technology has been employed for use in interspecies comparisons. In a comparison of the human genome against four great ape genomes, using a LIC array of 2460 BACs, 63 sites of DNA copy number variation between the human and great apes were identified. A significant number of these sites existed in interstitial euchromatin.<sup>56</sup> In a recent study, using a cDNA array CGH approach, over 29 000 human genes across five hominoid species (human, bonobo, chimpanzee, gorilla, and orangutan) were compared leading to the identification of >800 genes that gave genetic signatures unique to a specific hominoid lineage.<sup>57</sup> Moreover, there was a more pronounced difference between copy number increases and decreases in humans and a number of genes amplified are thought to be involved in the structure and function of the brain. These studies illustrate the use of array CGH in interspecies comparisons.

### Array CGH as a diagnostic tool and applications to clinical settings

The development of genomic and gene expression profiling technologies has allowed the simultaneous interrogation of thousands of loci and offers unprecedented opportunities to obtain global molecular signatures of the state of activity of cells in patient samples. The use of DNA-based technology has notable practical advantages in a clinical setting. DNA is stable, relatively easy to transport, and can be obtained from archival paraffin tissue blocks, while the acquisition and optimal transport of high-quality RNA is challenging due to its inherent instability. Furthermore, with array CGH, genomic DNA from normal cells of any origin from the same individual can be used as a baseline to define changes. In contrast, normal tissue from matching tissue type or precursor cells is required in order to properly define expression changes.

Current application of DNA-based diagnosis, primarily by standard cytogenetic analysis and chromosomal CGH, has wide application in a clinical setting, but suffers from low resolution and is not precisely linked to sequence-based map information. Array CGH offers the opportunity to globally profile segmental copy number imbalances at unprecedented resolution in constitutional or tumor DNA samples, thus serving as a diagnostic and investigative tool.

Disease-specific arrays have been constructed for cancer diagnostics. These arrays are enriched for the coverage of multiple cancer gene loci facilitating simultaneous assessment of gains and losses of tumor suppressor and oncogenes in a variety of cancers.<sup>12,42,43</sup> Similarly, diagnostic arrays have been designed for the diagnosis of congenital anomalies, developmental delay, and mental retardation<sup>58–60</sup> as well as the detection of chromosomal aberrations in embryos.<sup>49,61,62</sup> In order for array CGH to have a more prominent role in clinical diagnosis, many

factors such as cost, standardization of protocol, robustness of arrays, and user acceptance need to be addressed. Furthermore, advancements in array CGH software pertaining to ease of use, interpretation, visualization, and functionality will also be necessary.

### Array CGH data computational visualization and analysis

With the increase in array CGH applications and the diversity of platform technologies, a variety of software has been developed for data visualization and statistical analysis<sup>63–75</sup> (Table 1).

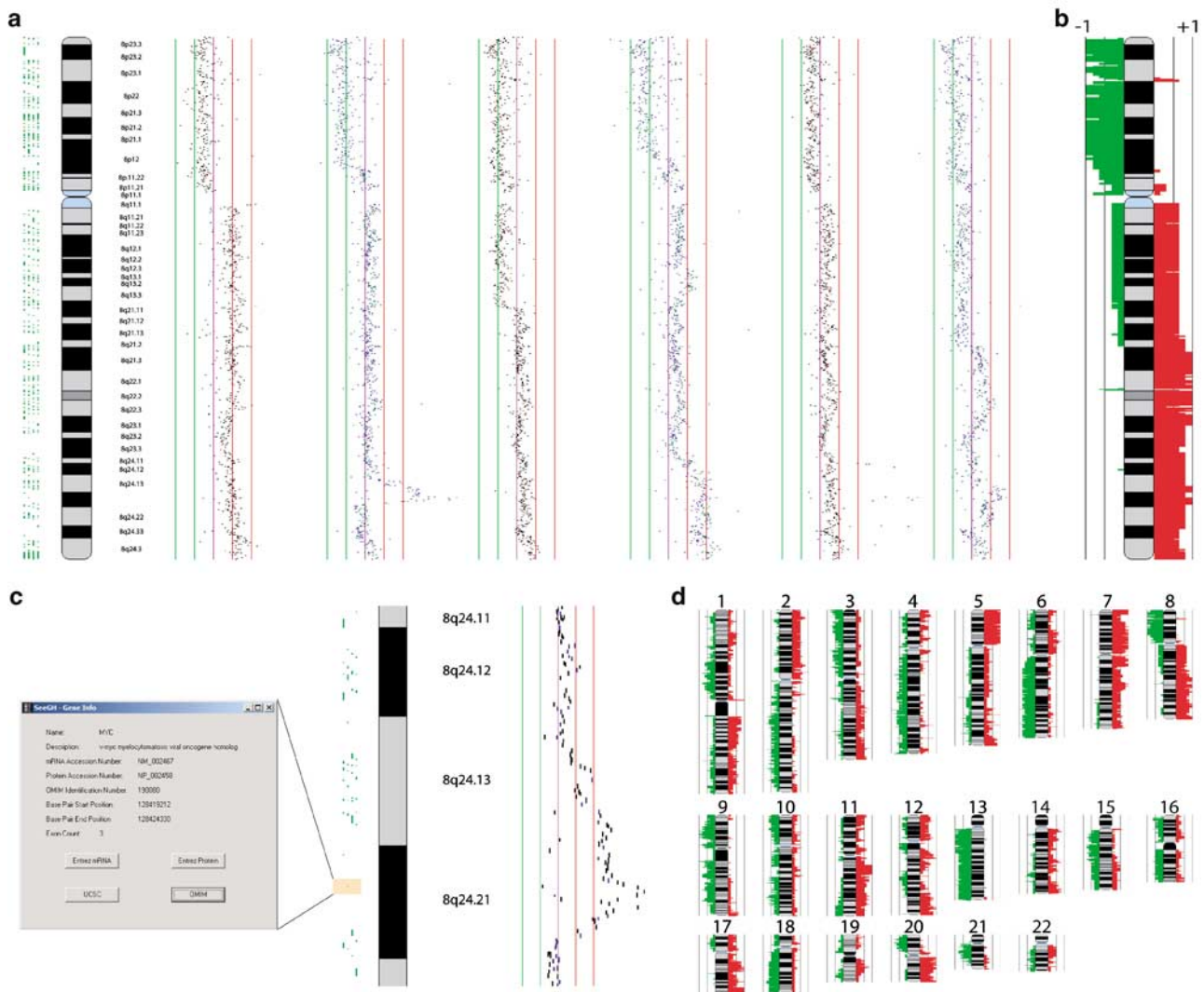
The first step in visualization is the conversion of spot image data to locus-specific copy number ratio, a function included in many microarray scanner software packages as well as custom software.<sup>76</sup> The next step involves the linking of the array elements to their genomic positions. The addressing is achieved by relating sequence information, such as the sequence of an oligonucleotide element or the end sequences of a BAC clone, to the human genome sequence, so that the signal ratio for each locus can be discretely displayed. This task becomes a challenge when displaying ratio data from tiling path arrays, where the array elements represent overlapping genomic segments.<sup>75</sup>

The identification and detection of segmental losses and gains requires statistical analysis. We have compared 17 publicly accessible software in terms of functionality, hardware and software requirements, input format, types of algorithms used, cost, and availability. We also compared the types of analysis supported: filtering and excluding data using cutoffs or thresholds, visualizing data at various levels of magnification, and viewing of multiple experiments simultaneously. Computationally basic software has the ability to view single experiments and to perform simple data analysis. This category of programs includes *ArrayCyCHt*, *Caryoscope*, and *SeeGH* v1.5. Computationally advanced software can view multiple experiments simultaneously, provide numerous forms of visualization, and perform sophisticated analysis techniques in determining alterations. Examples of computationally advanced software are *CGHPro*, *CGHAnalyzer* and *SeeGH* v2.0. Figure 4 illustrates some features of *SeeGH* v2.0 software including the alignment of multiple experiments and the calculation of frequency of alteration for array loci that are useful in the analysis of large data sets. A comparison of all the software programs is provided in Table 1. Although some of the software described can perform multiple types of analyses, further development is necessary to assemble functionalities in a single software package, that can support spot data normalization, map position addressing, multiple profile alignment, automated breakpoint detection, frequency, and cluster analysis, gene track referencing and alignment with exogenous data tracks such as complementary data on gene expression profiles and allelic status. Furthermore,

**Table 1** Software for array CGH visualization and analysis

Software	Cost	<sup>a</sup> Platform	Requirements	<sup>b</sup> Functions	Data Preprocessing	Breakpoint detection	Data input format	Profile Display	Web site and Reference
<i>aCGH analyzer v2.0</i>	Free	W	R V1.7.1 and Perl	V, S	Lowess normalization	No	Gene Pix (.gpr), text files	Single	genomic.dfc.harvard.edu/array_cgh_tools.asp
<i>aCGH smooth</i>	Free	W	Excel 97	V, S	Smoothing algorithm with modifiable parameters	Heuristic algorithm, regularized maximum likelihood, Threshold	Excel files	Single	www.few.vu.nl/~vumarray/Jong <i>et al</i> <sup>63</sup>
<i>Array CyGHt</i>	Free	Web-Based	None	V, S	Print-tip loess, global loess, robust spline normalization	Threshold	Text files	Single	genomics.catholic.ac.kr/arrayCGH/Kim <i>et al</i> <sup>64</sup>
<i>Caryoscope v0.3.10</i>	Free	W, M, L, U or Web-Based	JRE V1.4.2_03	V, S	No	Moving average, compute log ratio to any base	Text files, .gff files	Single	caryoscope.stanford.edu/Awad <i>et al</i> <sup>65</sup>
<i>CGH Analyzer</i>	Free	W, M, U	JRE V1.4.1	V, S	No	Standard ratio threshold, P-value based on reference	GenePix (.gpr), Affymetrix p501 SNP, text files	Multiple	guanine.genomics.upenn.edu/people/faculty/weberb/CGH/html/downloads.htm Margolin <i>et al</i> <sup>66</sup>
<i>CGHPRO</i>	Free	W, L	JRE V1.4.2, R and MySQL	V, S	Global Median, LOWESS, Subgrid Median, Subgrid LOWESS	Unsupervised Hidden Markov Partition, Circular Binary Segmentation	GenePix (.gpr), Agilent, Imagene, Text files	Multiple	molgen.mpg.de/~abt_rop/molecular_cytogenetics/ArrayCGH/CGHPRO Chen <i>et al</i> <sup>67</sup>
<i>CGH-Explorer</i>	Free	W, M, L	JRE V1.4	V, S	Centering, various data transformations, smoothing	Thresholding, bootstrap-based method, Analysis of Copy Errors (ACE)	Text files, Excel files	Multiple	www.ifi.uio.no/bioinf/Papers/CGH/Linjjaerde <i>et al</i> <sup>68</sup>
<i>CGH-Miner</i>	Free	W, U	Excel 2000, R	V, S	Simple (normalize each array to mean 0), smoothing	CLAC (clustering along chromosomes) with FDR (false discovery rate)	Excel files	Single (consensus plot of all expts)	www-stat.stanford.edu/%7Ewp57/CGH-Miner/Wang <i>et al</i> <sup>69</sup>
<i>CGH-Plotter</i>	Free	W, M, L	Matlab 6.1	V, S	Moving median, moving average	k-means clustering, dynamic programming	Text files, Matlab files (.mat)	Multiple	sigwww.cs.tut.fi/TICSP/CGH-Plotter Autio <i>et al</i> <sup>70</sup>
<i>CGH Analytics</i>	Cost, Free trial	W, M, L	None	V, S	User-selectable experimental calibration	Z-scoring, moving average calculation	Text files	Multiple	www.chem.agilent.com/Scripts/PDS.asp?Page=29457
<i>ChARM 1.7</i>	Free	W, M, L	JRE V1.4.2	V, S	Median filtering	Expectation Maximization (EM), one-sided sign test and/or mean permutation test	Text files, .pcl files	Multiple	function.princeton.edu/ChARM/Myers <i>et al</i> <sup>71</sup>
<i>DIGMAP Viewer 1.15</i>	Free	W	JRE V1.3	V, S	Compute log2 or log10 ratios, if not done so already	Clustering of cDNA expression data based on chromosome location	Text files, .csv files	Multiple	geneexplorer.mc.vanderbilt.edu/digmap/Yi <i>et al</i> <sup>72</sup>
<i>M-CGH</i>	Free	W, M, L, U	Matlab 6.1	V, S	Simple, intensity dependent or sub-array position normalization	Maximum likelihood and K-nearest neighbour or wavelet approach	GenePix (.gpr) or QuantArray files	Single	folk.uio.no/junbaiw/mcgh/Wang <i>et al</i> <sup>73</sup>
<i>Normalise Suite v2.5</i>	Free	W	None	V, S	Normalization, Smoothing	Region detection by user-defined thresholds or sliding window algorithm	GenePix (.gpr) files, Text files	Multiple	www.utoronto.ca/cancyto/index.html?protocols_software/software/index.html Beheshti <i>et al</i> <sup>74</sup>
<i>SeeGH v1.5</i>	Free	W	MySQL	V	N/A	N/A	Text files	Single	www.bccrc.ca/arraycgh/SeeGH.htm Chi <i>et al</i> <sup>75</sup>
<i>SeeGH v2.0</i>	License	W	MySQL	V, S	No	Moving average	Text files	Multiple	www.ArrayCGH.ca
<i>Spectral Ware 2.2</i>	Cost	Web-Based	JRE 1.4.2_05 plugin	V, S	11 algorithms for normalization, smoothing	Confidence interval, based on iterative algorithm	GenePix (.gpr), Imagene files	Single	www.spectralgenomics.com/spectralware.htm

<sup>a</sup>W = Windows; M = Macintosh; L = Linux; U = Unix.<sup>b</sup>V = visualization tool; S = statistical analysis.



**Figure 4** Analysis of array CGH data using *SeeGH* software. (a) Shows a multiple alignment of chromosome 8 from six tumor profiles. The main function of array CGH software packages is to link the array elements to genomic position. The elements are mapped according to base pair position to a specific chromosomal location. Segmental gains and losses are identified by shifting signal ratio to the left (loss) and right (gain) of the purple line, which represents a  $\log_2$  signal ratio of zero. (b) Illustrates a frequency plot summarizing genetic alterations. The vertical lines on each side of the chromosome represent the proportion of the samples containing a loss (green) or gain (red) of a particular array element. (c) Representation of the gene track (green lines) corresponding to the BAC clones (black lines) facilitating the linkage to public databases such as OMIM, NCBI Entrez and UCSC Genome Browser. For example, the amplification at 8q24.21 corresponds to the *c-Myc* oncogene. (d) A summary of the frequency of gains and losses in the six karyograms.

specialized software and user interface may have to be developed tailoring to clinical application as compared to research use.

## Conclusion

Array CGH technology has advanced greatly in the past decade with the development of a myriad of array platforms expanding its application to many aspects of genetic research.<sup>38,42,49,77</sup> The relative stability of DNA (as compared to RNA) and the ease of isolation allow

investigation of clinical specimens that may not be suitable for gene expression profiling. New array design will continue to improve resolution and detection sensitivity, while more efficient production strategies and streamlined experimental protocols will reduce cost and effort requirement. In addition, the emergence of new software aiming at automating breakpoint detection and statistical analysis will simplify the daunting task of the interpretation of array CGH data sets. The continuing technical advances and growing databases of disease-specific profiles will broaden the use of array CGH in both research and clinical settings.

## Acknowledgements

We thank B Coe, J Davies and T Buys for critical discussion and apologize for our inability to include all array CGH studies in this review due to size constraints. This work was supported by funds from Genome Canada, NIDCR ROI DE 015965-01, Lymphoma Research Foundation, CIHR, and NSERC of Canada scholarship to WWL.

## References

- 1 Solinas-Toldo S, Lampel S, Stilgenbauer S *et al*: Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* 1997; **20**: 399–407.
- 2 Pinkel D, Segraves R, Sudar D *et al*: High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* 1998; **20**: 207–211.
- 3 Inazawa J, Inoue J, Imoto I: Comparative genomic hybridization (CGH)-arrays pave the way for identification of novel cancer-related genes. *Cancer Sci* 2004; **95**: 559–563.
- 4 Oostlander AE, Meijer GA, Ylstra B: Microarray-based comparative genomic hybridization and its applications in human genetics. *Clin Genet* 2004; **66**: 488–495.
- 5 Pollack JR, Perou CM, Alizadeh AA *et al*: Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* 1999; **23**: 41–46.
- 6 Ishkanian AS, Malloff CA, Watson SK *et al*: A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* 2004; **36**: 299–303.
- 7 Pollack JR, Sorlie T, Perou CM *et al*: 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* 2002; **99**: 12963–12968.
- 8 Dhami P, Coffey AJ, Abbs S *et al*: Exon array CGH: detection of copy-number changes at the resolution of individual exons in the human genome. *Am J Hum Genet* 2005; **76**: 750–762.
- 9 Snijders AM, Nowak N, Segraves R *et al*: Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 2001; **29**: 263–264.
- 10 Vissers LE, de Vries BB, Osoegawa K *et al*: Array-based comparative genomic hybridization for the genomewide detection of sub-microscopic chromosomal abnormalities. *Am J Hum Genet* 2003; **73**: 1261–1270.
- 11 Fiegler H, Carr P, Douglas EJ *et al*: DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* 2003; **36**: 361–374.
- 12 Greshock J, Naylor TL, Margolin A *et al*: 1-Mb resolution array-based comparative genomic hybridization using a BAC clone set optimized for cancer gene analysis. *Genome Res* 2004; **14**: 179–187.
- 13 Snijders AM, Pinkel D, Albertson DG: Current status and future prospects of array-based comparative genomic hybridization. *Brief Funct Genomic Proteomic* 2003; **2**: 37–45.
- 14 Bignell GR, Huang J, Greshock J *et al*: High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* 2004; **14**: 287–295.
- 15 Kennedy GC, Matsuzaki H, Dong S *et al*: Large-scale genotyping of complex DNA. *Nat Biotechnol* 2003; **21**: 1233–1237.
- 16 Zhao X, Li C, Paez JG *et al*: An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res* 2004; **64**: 3060–3071.
- 17 Matsuzaki H, Dong S, Loi H *et al*: Genotyping over 100 000 SNPs on a pair of oligonucleotide arrays. *Nat Meth* 2004; **1**: 109–111.
- 18 Brennan C, Zhang Y, Leo C *et al*: High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res* 2004; **64**: 4744–4748.
- 19 Carvalho B, Ouwerkerk E, Meijer GA, Ylstra B: High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J Clin Pathol* 2004; **57**: 644–646.
- 20 Barrett MT, Scheffer A, Ben-Dor A *et al*: Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc Natl Acad Sci USA* 2004; **101**: 17765–17770.
- 21 Selzer RR, Richmond TA, Pofahl NJ *et al*: Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosomes Cancer* 2005; **44**: 305–319.
- 22 Lucito R, Healy J, Alexander J *et al*: Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res* 2003; **13**: 2291–2305.
- 23 Wang TL, Maierhofer C, Speicher MR *et al*: Digital karyotyping. *Proc Natl Acad Sci USA* 2002; **99**: 16156–16161.
- 24 Tuzun E, Sharp AJ, Bailey JA *et al*: Fine-scale structural variation of the human genome. *Nat Genet* 2005; **37**: 727–732.
- 25 Squire JA, Pei J, Marrano P *et al*: High-resolution mapping of amplifications and deletions in pediatric osteosarcoma by use of CGH analysis of cDNA microarrays. *Genes Chromosomes Cancer* 2003; **38**: 215–225.
- 26 Pettus JA, Cowley BC, Maxwell T *et al*: Multiple abnormalities detected by dye reversal genomic microarrays in prostate cancer: a much greater sensitivity than conventional cytogenetics. *Cancer Genet Cytogenet* 2004; **154**: 110–118.
- 27 Garnis C, Coe BP, Lam SL, Macaulay C, Lam WL: High-resolution array CGH increases heterogeneity tolerance in the analysis of clinical samples. *Genomics* 2005; **85**: 790–793.
- 28 Coe BP, Henderson LJ, Garnis C *et al*: High-resolution chromosome arm 5p array CGH analysis of small cell lung carcinoma cell lines. *Genes Chromosomes Cancer* 2005; **42**: 308–313.
- 29 Garnis C, Baldwin C, Zhang L, Rosin MP, Lam WL: Use of complete coverage array comparative genomic hybridization to define copy number alterations on chromosome 3p in oral squamous cell carcinomas. *Cancer Res* 2003; **63**: 8582–8585.
- 30 Henderson LJ, Coe BP, Lee EH *et al*: Genomic and gene expression profiling of minute alterations of chromosome arm 1p in small-cell lung carcinoma cells. *Br J Cancer* 2005; **92**: 1553–1560.
- 31 Garnis C, Davies JJ, Buys TP *et al*: Chromosome 5p aberrations are early events in lung cancer: implication of glial cell line-derived neurotrophic factor in disease progression. *Oncogene* 2005; **24**: 4806–4812.
- 32 Garnis C, Campbell J, Davies JJ, Macaulay C, Lam S, Lam WL: Involvement of multiple developmental genes on chromosome 1p in lung tumorigenesis. *Hum Mol Genet* 2005; **14**: 475–482.
- 33 Seng TJ, Ichimura K, Liu L, Tingby O, Pearson DM, Collins VP: Complex chromosome 22 rearrangements in astrocytic tumors identified using microsatellite and chromosome 22 tile path array analysis. *Genes Chromosomes Cancer* 2005; **43**: 181–193.
- 34 Buckley PG, Mantripragada KK, Benetkiewicz M *et al*: A full-coverage, high-resolution human chromosome 22 genomic microarray for clinical and research applications. *Hum Mol Genet* 2002; **11**: 3221–3229.
- 35 Holzmann K, Kohlhammer H, Schwaenen C *et al*: 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* 2004; **64**: 4428–4433.
- 36 Zielinski B, Gratias S, Toedt G *et al*: Detection of chromosomal imbalances in retinoblastoma by matrix-based comparative genomic hybridization. *Genes Chromosomes Cancer* 2005; **43**: 294–301.
- 37 Takada H, Imoto I, Tsuda H *et al*: Screening of DNA copy-number aberrations in gastric cancer cell lines by array-based comparative genomic hybridization. *Cancer Sci* 2005; **96**: 100–110.
- 38 Baldwin C, Garnis C, Zhang L, Rosin MP, Lam WL: Multiple microalterations detected at high frequency in oral cancer. *Cancer Res* 2005; **65**: 7561–7567.

- 39 van Dekken H, Paris PL, Albertson DG *et al*: Evaluation of genetic patterns in different tumor areas of intermediate-grade prostatic adenocarcinomas by high-resolution genomic array analysis. *Genes Chromosomes Cancer* 2004; **39**: 249–256.
- 40 Rubio-Moscardo F, Climent J, Siebert R *et al*: Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. *Blood* 2005; **105**: 4445–4454.
- 41 Pang E, Hu Y, Chan KY *et al*: Karyotypic imbalances and differential gene expressions in the acquired doxorubicin resistance of hepatocellular carcinoma cells. *Lab Invest* 2005; **85**: 664–674.
- 42 Schwaenen C, Nessling M, Wessendorf S *et al*: 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* 2004; **101**: 1039–1044.
- 43 Kohlhammer H, Schwaenen C, Wessendorf S *et al*: 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* 2004; **104**: 795–801.
- 44 de Leeuw RJ, Davies JJ, Rosenwald A *et al*: Comprehensive whole genome array CGH profiling of mantle cell lymphoma model genomes. *Hum Mol Genet* 2004; **13**: 1827–1837.
- 45 Lim G, Karaskova J, Beheshti B *et al*: An integrated mBAND and submegabase resolution tiling set (SMRT) CGH array analysis of focal amplification, microdeletions, and ladder structures consistent with breakage-fusion-bridge cycle events in osteosarcoma. *Genes Chromosomes Cancer* 2005; **42**: 392–403.
- 46 Coe BP, Lee EH, Chi B *et al*: Tiling genomic array profiling of small cell lung cancer cell lines identifies highly frequent copy number gains of MAD1L1. *Genes Chromosomes Cancer* 2005; **45**: 11–19.
- 47 Garnis C, Lockwood WW, Vucic E *et al*: Comparing canonical non-small cell lung squamous and adenocarcinoma model genomes. *Int J Cancer* 2005; In Press.
- 48 Shaw-Smith C, Redon R, Rickman L *et al*: Microarray based comparative genomic hybridisation (array-CGH) detects submicroscopic chromosomal deletions and duplications in patients with learning disability/mental retardation and dysmorphic features. *J Med Genet* 2004; **41**: 241–248.
- 49 Schaeffer AJ, Chung J, Heretis K, Wong A, Ledbetter DH, Lese Martin C: Comparative genomic hybridization-array analysis enhances the detection of aneuploidies and submicroscopic imbalances in spontaneous miscarriages. *Am J Hum Genet* 2004; **74**: 1168–1174.
- 50 Klaassens M, van Dooren M, Eussen HJ *et al*: Congenital diaphragmatic hernia and chromosome 15q26: determination of a candidate region by use of fluorescent *in situ* hybridization and array-based comparative genomic hybridization. *Am J Hum Genet* 2005; **76**: 877–882.
- 51 Klein OD, Cotter PD, Albertson DG *et al*: Prader–Willi syndrome resulting from an unbalanced translocation: characterization by array comparative genomic hybridization. *Clin Genet* 2004; **65**: 477–482.
- 52 Zhang X, Snijders A, Seagraves R *et al*: High-resolution mapping of genotype-phenotype relationships in cri-du-chat syndrome using array comparative genomic hybridization. *Am J Hum Genet* 2005; **76**: 312–326.
- 53 Iafrate AJ, Feuk L, Rivera MN *et al*: Detection of large-scale variation in the human genome. *Nat Genet* 2004; **36**: 949–951.
- 54 Sharp AJ, Locke DP, McGrath SD *et al*: Segmental duplications and copy-number variation in the human genome. *Am J Hum Genet* 2005; **77**: 78–88.
- 55 Sebat J, Lakshmi B, Troge J *et al*: Large-scale copy number polymorphism in the human genome. *Science* 2004; **305**: 525–528.
- 56 Locke DP, Seagraves R, Carbone L *et al*: Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. *Genome Res* 2003; **13**: 347–357.
- 57 Fortna A, Kim Y, MacLaren E *et al*: Lineage-specific gene duplication and loss in human and great ape evolution. *PLoS Biol* 2004; **2**: E207.
- 58 Bejjani BA, Saleki R, Ballif BC *et al*: Use of targeted array-based CGH for the clinical diagnosis of chromosomal imbalance: is less more? *Am J Med Genet A* 2005; **134**: 259–267.
- 59 Yu W, Ballif BC, Kashork CD *et al*: Development of a comparative genomic hybridization microarray and demonstration of its utility with 25 well-characterized 1p36 deletions. *Hum Mol Genet* 2003; **12**: 2145–2152.
- 60 Cheung SW, Shaw CA, Yu W *et al*: Development and validation of a CGH microarray for clinical cytogenetic diagnosis. *Genet Med* 2005; **7**: 422–432.
- 61 Wilton L, Williamson R, McBain J, Edgar D, Voullaire L: Birth of a healthy infant after preimplantation confirmation of euploidy by comparative genomic hybridization. *N Engl J Med* 2001; **345**: 1537–1541.
- 62 Larrabee PB, Johnson KL, Pestova E *et al*: Microarray analysis of cell-free fetal DNA in amniotic fluid: a prenatal molecular karyotype. *Am J Hum Genet* 2004; **75**: 485–491.
- 63 Jong K, Marchiori E, Meijer G, Vaart AV, Ylstra B: Breakpoint identification and smoothing of array comparative genomic hybridization data. *Bioinformatics* 2004; **20**: 3636–3637.
- 64 Kim SY, Nam SW, Lee SH *et al*: ArrayCyGH: a web application for analysis and visualization of array-CGH data. *Bioinformatics* 2005; **21**: 2554–2555.
- 65 Awad IA, Rees CA, Hernandez-Boussard T, Ball CA, Sherlock G: Caryoscope: an Open Source Java application for viewing microarray data in a genomic context. *BMC Bioinformatics* 2004; **5**: 151.
- 66 Margolin AA, Greshock J, Naylor TL *et al*: CGHAnalyzer: a stand-alone software package for cancer genome analysis using array-based DNA copy number data. *Bioinformatics* 2005; **21**: 3308–3311.
- 67 Chen W, Erdogan F, Ropers HH, Lenzner S, Ullmann R: CGHPRO – a comprehensive data analysis tool for array CGH. *BMC Bioinformatics* 2005; **6**: 85.
- 68 Lingjaerde OC, Baumbusch LO, Liestol K, Glad IK, Borresen-Dale AL: CGH-Explorer: a program for analysis of array-CGH data. *Bioinformatics* 2005; **21**: 821–822.
- 69 Wang P, Kim Y, Pollack J, Narasimhan B, Tibshirani R: A method for calling gains and losses in array CGH data. *Biostatistics* 2005; **6**: 45–58.
- 70 Autio R, Hautaniemi S, Kauraniemi P *et al*: CGH-Plotter: MATLAB toolbox for CGH-data analysis. *Bioinformatics* 2003; **19**: 1714–1715.
- 71 Myers CL, Dunham MJ, Kung SY, Troyanskaya OG: Accurate detection of aneuploidies in array CGH and gene expression microarray data. *Bioinformatics* 2004; **20**: 3533–3543.
- 72 Yi Y, Mirosevich J, Shyr Y, Matusik R, George Jr AL: Coupled analysis of gene expression and chromosomal location. *Genomics* 2005; **85**: 401–412.
- 73 Wang J, Meza-Zepeda LA, Kresse SH, Myklebost O: M-CGH: analysing microarray-based CGH experiments. *BMC Bioinformatics* 2004; **5**: 74.
- 74 Beheshti B, Braude I, Marrano P, Thorner P, Zielenska M, Squire JA: Chromosomal localization of DNA amplifications in neuroblastoma tumors using cDNA microarray comparative genomic hybridization. *Neoplasia* 2003; **5**: 53–62.
- 75 Chi B, DeLeeuw RJ, Coe BP, MacAulay C, Lam WL: SeeGH – a software tool for visualization of whole genome array comparative genomic hybridization data. *BMC Bioinformatics* 2004; **5**: 13.
- 76 Jain AN, Tokuyasu TA, Snijders AM, Seagraves R, Albertson DG, Pinkel D: Fully automatic quantification of microarray image data. *Genome Res* 2002; **12**: 325–332.
- 77 Weber M, Davies JJ, Wittig D *et al*: Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005; **37**: 853–862.

Supplementary Information accompanies the paper on European Journal of Human Genetic website (<http://www.nature.com/ejhg>)