

A tiling resolution DNA microarray with complete coverage of the human genome

Adrian S Ishkanian¹, Chad A Malloff¹, Spencer K Watson¹, Ronald J deLeeuw¹, Bryan Chi¹, Bradley P Coe¹, Antoine Snijders², Donna G Albertson², Daniel Pinkel², Marco A Marra³, Victor Ling¹, Calum MacAulay¹ & Wan L Lam¹

We constructed a tiling resolution array consisting of 32,433 overlapping BAC clones covering the entire human genome. This increases our ability to identify genetic alterations and their boundaries throughout the genome in a single comparative genomic hybridization (CGH) experiment. At this tiling resolution, we identified minute DNA alterations not previously reported. These alterations include microamplifications and deletions containing oncogenes, tumor-suppressor genes and new genes that may be associated with multiple tumor types. Our findings show the need to move beyond conventional marker-based genome comparison approaches, that rely on inference of continuity between interval markers. Our submegabase resolution tiling set for array CGH (SMRT array) allows comprehensive assessment of genomic integrity and thereby the identification of new genes associated with disease.

Identification of chromosomal imbalances and variation in DNA copy-number is essential to our understanding of disease mechanisms and pathogenesis. Array CGH¹ or matrix CGH² offers the highest resolution for practical genome-wide detection of chromosomal alterations. This technique is derived from the concept of conventional CGH³, which has contributed greatly to the molecular characterization of both somatic and constitutional genomic DNA mutations over the last decade^{4–6}. The primary limitation of conventional CGH is its resolution (~20 Mb), as this method detects segmental copy-number changes on metaphase chromosomes³. In array CGH, the metaphase chromosome spread is replaced by BACs, PACs or YACs containing human DNA as targets, increasing the resolution to the distance between the selected marker DNA clones^{1,2}. Genome screening using array CGH has great potential in the characterization of numerous chromosomal disorders.

Efforts to construct DNA arrays spanning the human genome consisted of spotting 2,460 (ref. 7) or 3,500 (ref. 8) marker BAC clones representing the sequenced genome at an average interval of ~1 Mb.

These studies showed that sufficient target-DNA printing solution could be generated from individual BACs using PCR-based protocols. Because the target product is PCR-derived, it is easily replenishable, obviating the need for multiple rounds of laborious large-scale BAC DNA preparations. These arrays are sensitive enough to detect single-copy changes, but the technique is limited by the small number of BAC markers representing the genome on the slide, rather than the methodology. Even at this resolution, array CGH is useful for detecting chromosomal aberrations associated with congenital abnormalities and somatic malignancies^{9–12}.

Recent studies focused on higher-density regional arrays for fine mapping and identifying new genes in specific chromosomal regions^{13–18}. For example, a candidate oncogene for association with

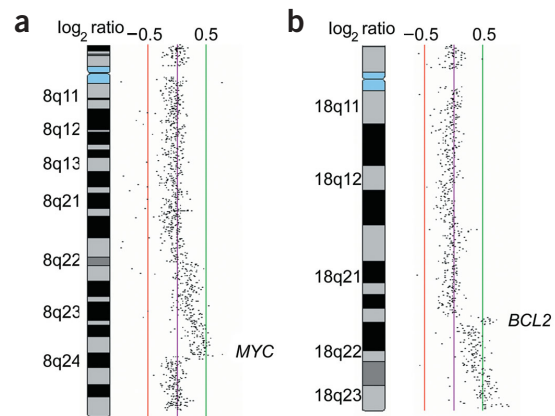


Figure 1 Detection of twofold copy-number changes in TAT-1 lymphoma cell line on chromosome arms 8q and 18q. **(a)** Chromosome view of 8q showing *MYC* amplification between BAC clones RP11-143H8 and RP11-263C20. **(b)** Chromosome view of 18q showing *BCL2* amplification between BAC clones RP11-159K14 and RP11-565D23. Vertical green and red lines are scale bars indicating \log_2 ratios of +0.5 and -0.5, respectively.

¹British Columbia Cancer Research Centre, 601 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada. ²UCSF Comprehensive Cancer Center, 2340 Sutter Street, San Francisco, California 94115, USA. ³Genome Sciences Centre, British Columbia Cancer Agency, 600 West 10th Avenue, Vancouver BC V5Z 4E6, Canada. Correspondence should be addressed to W.L.L. (wlanlam@bccrc.ca).

breast cancer (*CYP24A1*) was identified on 20q13.2 using an array of 29 overlapping clones spanning this region¹⁸. The need for a tiling resolution array to map these amplification or deletion boundaries is indicated by the fact that two separate regions of amplification in 20q13.2 contained two separate putative oncogenes, which would not have been detected by a lower resolution array. These studies show that the resolving power of array CGH is maximized when the detection of single copy-number changes is combined with a tiling or overlapping set of BAC clones.

We created the first tiling resolution BAC array with complete coverage of the human genome using 32,433 fingerprint-verified individually amplified BAC clones. Here we show that such a complete genome comparison is capable of identifying microamplifications and

microdeletions, which may contain genes involved in disease pathogenesis. We call this array submegabase resolution tiling set for array CGH (SMRT array).

RESULTS

Array sensitivity

To assess the sensitivity of the SMRT array, we hybridized the well-characterized EBV-transformed lymphoma cell line TAT-1 (ref. 19) to normal male genomic DNA. Genomic regions containing *BCL2* (18q21) and *MYC* (8q24) in TAT-1 were previously shown to have a twofold copy-number increase by FISH analysis¹⁹. We detected these previously reported amplifications at both loci, and we delineated their boundaries (Fig. 1). Boundaries of amplification on

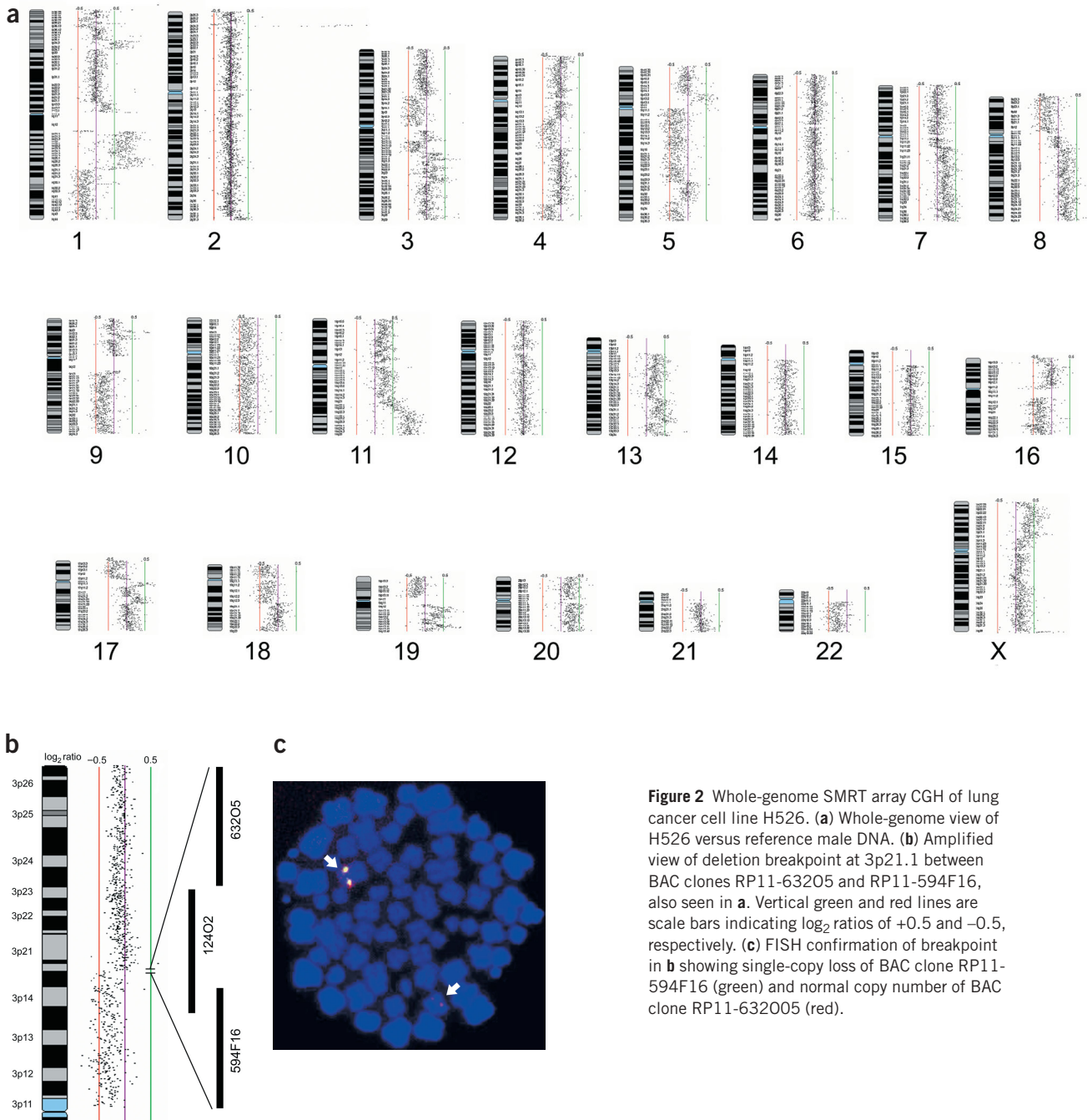


Figure 2 Whole-genome SMRT array CGH of lung cancer cell line H526. (a) Whole-genome view of H526 versus reference male DNA. (b) Amplified view of deletion breakpoint at 3p21.1 between BAC clones RP11-63205 and RP11-594F16, also seen in a. Vertical green and red lines are scale bars indicating log₂ ratios of +0.5 and -0.5, respectively. (c) FISH confirmation of breakpoint in b showing single-copy loss of BAC clone RP11-594F16 (green) and normal copy number of BAC clone RP11-632005 (red).



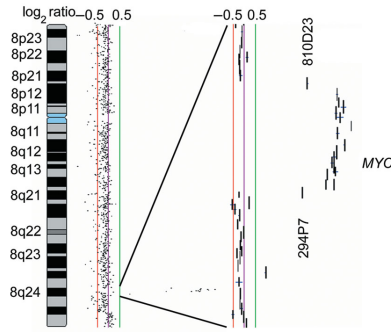


Figure 3 Amplification of chromosome 8q24.12–13 in colorectal cancer cell line COLO320. This 1.9-Mb amplification containing *MYC* is bounded by BAC clones RP11-810D23 and RP11-294P7. Vertical green and red lines are scale bars indicating \log_2 ratios of +0.5 and -0.5, respectively.

chromosome 8 were between BAC clone RP11-143H8 at 8q22.2 and RP11-263C20 at 8q24.13. Boundaries of amplification on chromosome 18 were between BAC clone RP11-159K14 at 18q21.32 and RP11-565D23 at 18q23. These data illustrate the detection sensitivity of array CGH.

Array resolution compared to conventional CGH

To demonstrate the resolving power of the SMRT array, we compared the \log_2 ratio profile of lung cancer cell line H526 (refs. 20,21; Fig. 2a) to the previously published conventional chromosomal CGH data (see URLs). All patterns of gains and losses were matched, including large changes (e.g., the amplification of 7q and 8q and loss of the entire chromosome 10) and complex changes (e.g., the multiple amplifications on chromosome 1 and the multiple deletions on chromosome 4). Notably, conventional chromosomal CGH identified a highly amplified region on the telomeric end of chromosome arm 2p, apparently covering approximately one-fourth of the whole chromosome. However, the SMRT array analysis showed this amplification to be precisely localized to a 1.3-Mb fragment at 2p24.3, bordered by BAC clones RP11-351F4 and RP11-701O10, which contains the *MYCN* oncogene. The resolving power of this whole-genome array enables us to define breakpoints to within single BAC clones. For example, the deletion breakpoint on chromosome arm 3p was localized to between BAC clones RP11-632O5 and RP11-594F16 at 3p21.1 (Fig. 2b). This finding was subsequently confirmed by FISH analysis (Fig. 2c).

Comparison to previous array CGH

To compare our tiling resolution array against current array CGH technology, we profiled colorectal cancer cell line COLO320 (ref. 22), which has been characterized in two previous array CGH studies^{7,23}. We confirmed the amplification at 8q24 in the *MYC* region identified by these studies. Furthermore, the SMRT system further defined this segmental copy-number increase precisely to a 1.9-Mb region bordered by BAC clones RP11-810D23 and RP11-294P7 (Fig. 3).

A detailed analysis of our COLO320 profile identified new microamplifications on chromosome arms 13q, 15q, 16p and 22q (Supplementary Fig. 1 online), which were not detected by the two previous high-resolution CGH studies^{7,23}. For example, we identified a 300-kb microamplification at 13q12.2 containing only three genes (according to University of California Santa Cruz Genome Browser, April 2003 Freeze): caudal type homeobox transcription factor 2 (*CDX2*), insulin promoter factor 1 (*IPF1*) and GS homeobox 1 (*GSH1*;

Fig. 4a). *CDX2* is a transcription factor expressed in the intestine and altered in colorectal cancers²⁴. FISH analysis verified this microamplification and showed that it was within a homogeneously staining region (Fig. 4b). These findings illustrate the usefulness of a tiling resolution BAC array for comprehensive assessment of genomic integrity.

Identification of minute regions of alteration

In addition to microamplifications, we also detected small deletions in a number of tumor cell lines. For example, we detected a 1.25-Mb deletion containing the gene *CDKN2A* (also called *p16*) in lymphoma cell line Z138C at 9p21.3 (Fig. 5a). Deletion of *CDKN2A* occurs in approximately one-half of mantle cell lymphoma tumors as detected by FISH²⁵. This deletion is bordered by RP11-328C2 and RP11-275H17 (Fig. 5a). Submegabase-sized microdeletions can be accurately mapped in a single whole-genome array CGH experiment. This is made possible by the overlapping clone coverage and their distribution on the array. A notable example is a 240-kb deletion at 7q22.3 in the breast cancer cell line BT474, containing *PRKAR2B*, a regulatory kinase, and *HBPI*, a G1 inhibitory kinase regulated by p38 MAP kinase²⁶ (Fig. 5b). Such microdeletions have not been reported previously. The mechanisms by which such deletions are effected are not known. Whether this microdeletion affects the expression of *PRKAR2B* or the neighboring gene, *PIK3CG*, remains to be determined. The two experiments described here show how small, previously unidentified alterations that have the potential to contribute to disease may easily be identified in a single SMRT array experiment.

DISCUSSION

Array CGH is a proven method for accurate, robust and rapid genome-wide assessment of DNA copy-number variation. Current users of array CGH technology consider BAC DNA markers positioned at intervals of 1–2 Mb to be ‘high-resolution’ coverage. This view has been perpetuated by conventional whole-genome analysis tools, such as microsatellite marker analysis of loss of heterozygosity, in which small interspaced ‘sequence-tagged sites’ are assayed for

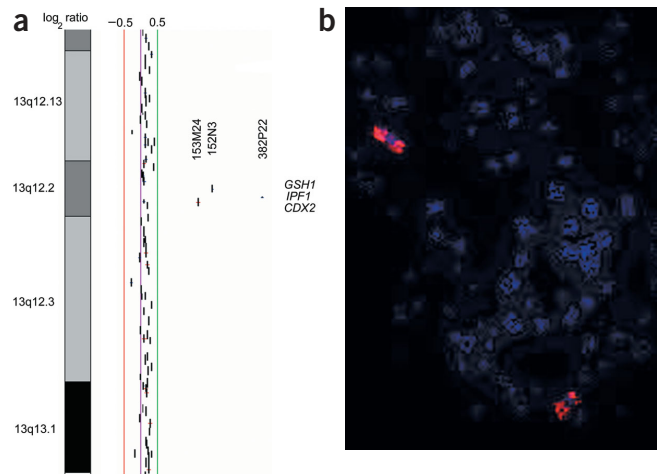


Figure 4 Identification of a new microamplification by tiling resolution array CGH in COLO320. (a) 300-kb microamplification on chromosome 13q12.2 containing genes *GSH1*, *CDX2* and *IPF1* and bounded by BAC clones RP11-153M24 and RP11-152N3. Vertical green and red lines are scale bars indicating \log_2 ratios of +0.5 and -0.5, respectively. (b) High copy-number amplification of RP11-153M24 detected by FISH hybridization. Amplification is located in a homogeneously staining region.

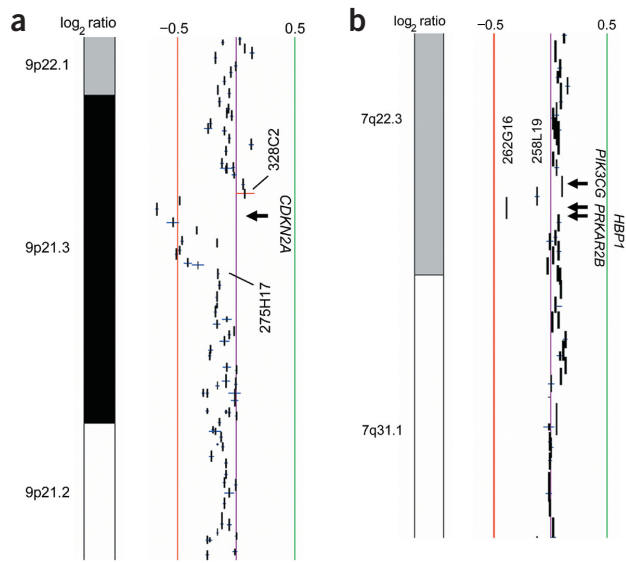


Figure 5 Identification of microdeletions. (a) Identification of a 1.25-Mb deletion at 9p21.3 in a mantle cell lymphoma cell line containing *CDKN2A* bounded by BAC clones RP11-328C2 and RP11-275H17. (b) 240-kb deletion at 7q22.3 in breast cancer cell line BT474 containing *PRKAR2B* and *HBP1* bounded by BAC clones RP11-258L19 and RP11-262G16. Vertical green and red lines are scale bars indicating \log_2 ratios of +0.5 and -0.5, respectively.

genomic imbalance, and the genomic integrity between these sites must be inferred. In contrast, tiling resolution array CGH has the potential to identify minute genomic changes. In this study, we constructed a submegabase resolution tiling set for array CGH (SMRT array), comprising 32,433 overlapping BAC clones covering the entire human genome. This tiling resolution, combined with the proven sensitivity of array CGH, makes the technique ideal for identifying new genes and will prove useful for unraveling the genetic basis of numerous diseases.

METHODS

BAC clone selection, preparation and validation. Selection and the map positions of the 32,433 clones has been described previously and is available from The Children's Hospital Oakland Research Institute (see URLs). We validated clone identity by comparing *Hind*III fingerprints to the FPC BAC fingerprint database²⁷ (see URLs). These clones provide ~1.5-fold coverage of the human genome, giving an approximate resolution of 80 kb (*i.e.*, two-thirds of an average BAC clone).

Array production from BAC DNA. We prepared the DNA samples to be spotted on the array by PCR using linkers (primer sequences available on request). The protocol for linker-mediated PCR was previously described²⁸. We precipitated the PCR products with ethanol and redissolved them in an MSP printing solution (Telechem), denatured them by boiling, and rearranged them for robotic printing in triplicate using a VersArray ChipWriter Pro (BioRad). This arrayer uses a 12 × 4 array of SMP2.5 Stealth Micro Spotting Pins (Telechem/ArrayIT) depositing DNA spots of 0.8 nl at ~1 μg μl⁻¹ at 133-μm distances. We spotted the entire set of 32,433 solutions in triplicate onto two aldehyde-coated slides. Limited numbers of SMRT arrays are available on a cost recovery basis (see URLs).

DNA labeling and hybridization. We labeled 400 ng of test and reference DNA separately with Cyanine-3 and Cyanine-5 dCTPs according to a random priming protocol previously described¹³. Before hybridization, we combined the

DNA probes and purified them using ProbeQuant Sephadex G-50 Columns (Amersham) to remove unincorporated nucleotides. We then added 200 μg of human Cot-1 DNA (Invitrogen), precipitated the mixture and resuspended it in 100 μl of DIG Easy hybridization solution (Roche) containing sheared herring sperm DNA (Sigma-Aldrich) and yeast tRNA (Calbiochem). The probe was denatured at 85 °C for 10 min and repetitive sequences were blocked at 45 °C for 1 h before hybridization. We carried out prehybridization in the same buffer. We applied the probe mixture to the slide surface, fixed the coverslips and incubated them at 42 °C for 36 h. We washed the arrays five times for 5 min each in 0.1× saline sodium citrate, 0.1% SDS at room temperature with agitation. We then rinsed each array repeatedly in 0.1× saline sodium citrate and dried them by centrifugation.

Array imaging and analysis. We imaged hybridized slides using a CCD-based imaging system (Arraywrx eAuto, Applied Precision) and analyzed them with SoftWoRx Tracker Spot Analysis software. We averaged the ratios of the triplicate spots and calculated standard deviations (*s.d.*). All spots with *s.d.* >0.075 or signal-to-noise ratios <20 were removed from the analysis. We used custom viewing software (SeeGH) to visualize all data as \log_2 ratio plots where each dot represents one BAC. This software is available on request (see URLs).

Reference male versus reference female hybridization detected no unexpected gains or losses, and random variability of \log_2 ratios were not observed (Supplementary Fig. 2 online). Furthermore, owing to overlapping clone coverage, a single clone with aberrant signal ratio would not be considered an amplification or deletion. Finally, since the clones are not spotted in the order of their map position, adjacent clones are distributed throughout our array.

URLs. The H526 CGH profile is available at <http://amba.charite.de/~ksch/cgh-database/index.htm>. The URL for Children's Hospital Oakland Research Institute is <http://bacpac.chori.org/genomicRearrays.php>. The FPC database is available at <http://genome.wustl.edu/projects/human/index.php?fpc=1>. Whole-genome DNA arrays and SeeGH software is available at http://www.bccrc.ca/cg/arraycgh_group.html.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank J. Vielkind, S. Lam, D. Horsman, M. Rosin, S. Herst, K. Lonergan, S. Ralph, J. Rathmann, R. Seagraves, M. Krzywinski, P. Lansdorp, G. Bebb, J. Schein, I. Bosdet, D. Smailus, Z. Xu, C. Brown, J. Minna and A. Gazdar for their contributions. This project was supported by funds from Genome Canada/British Columbia.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 16 December 2003; accepted 23 January 2004

Published online at <http://www.nature.com/naturegenetics/>

1. Pinkel, D. *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20**, 207–211 (1998).
2. Solinas-Toldo, S. *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**, 399–407 (1997).
3. Kallioniemi, A. *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258**, 818–821 (1992).
4. Knuutila, S., Autio, K. & Aalto, Y. Online access to CGH data of DNA sequence copy number changes. *Am. J. Pathol.* **157**, 689 (2000).
5. Forozan, F., Karhu, R., Kononen, J., Kallioniemi, A. & Kallioniemi, O.P. Genome screening by comparative genomic hybridization. *Trends Genet.* **13**, 405–409 (1997).
6. Wells, D. & Levy, B. Cytogenetics in reproductive medicine: the contribution of comparative genomic hybridization (CGH). *Bioessays* **25**, 289–300 (2003).
7. Snijders, A.M. *et al.* Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat. Genet.* **29**, 263–264 (2001).
8. Fiegler, H. *et al.* DNA microarrays for comparative genomic hybridization based on DOP-PCR amplification of BAC and PAC clones. *Genes Chromosomes Cancer* **36**, 361–374 (2003).
9. Veltman, J.A. *et al.* High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. *Am. J. Hum. Genet.* **70**, 1269–1276 (2002).
10. Kraus, J., Pantel, K., Pinkel, D., Albertson, D.G. & Speicher, M.R. High-resolution genomic profiling of occult micrometastatic tumor cells. *Genes Chromosomes Cancer* **36**, 159–166 (2003).



11. Veltman, J.A. *et al.* Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res.* **63**, 2872–2880 (2003).
12. Weiss, M.M. *et al.* Genomic profiling of gastric cancer predicts lymph node status and survival. *Oncogene* **22**, 1872–1879 (2003).
13. Garnis, C., Baldwin, C., Zhang, L., Rosin, M.P. & Lam, W.L. Use of complete coverage array CGH to define copy number alterations on chromosome 3p in oral squamous cell carcinomas. *Cancer Res.* **63**, 8582–8585 (2003).
14. Wilhelm, M. *et al.* Array-based comparative genomic hybridization for the differential diagnosis of renal cell cancer. *Cancer Res.* **62**, 957–960 (2002).
15. Bruder, C.E. *et al.* High resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. *Hum. Mol. Genet.* **10**, 271–282 (2001).
16. Garnis, C., Campbell, J., Zhang, L., Rosin, M.P. & Lam, W.L. OCGR Array, an oral cancer genomic regional array for comparative genomic hybridization analysis. *Oral Oncol.* (in the press).
17. Garnis, C. *et al.* 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 (2004).
18. Albertson, D.G. *et al.* Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. *Nat. Genet.* **25**, 144–146 (2000).
19. Denyssevych, T. *et al.* Establishment and comprehensive analysis of a new human transformed follicular lymphoma B cell line, Tat-1. *Leukemia* **16**, 276–283 (2002).
20. Levin, N.A. *et al.* Identification of frequent novel genetic alterations in small cell lung carcinoma. *Cancer Res.* **54**, 5086–5091 (1994).
21. Girard, L. *et al.* Genome-wide allelotyping of lung cancer identifies new regions of allelic loss, differences between small cell lung cancer and non-small cell lung cancer, and loci clustering. *Cancer Res.* **60**, 4894–4906 (2000).
22. Quinn, L.A., Moore, G.E., Morgan, R.T. & Woods, L.K. Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions. *Cancer Res.* **39**, 4914–4924 (1979).
23. Wessendorf, S. *et al.* Automated screening for genomic imbalances using matrix-based comparative genomic hybridization. *Lab. Invest.* **82**, 47–60 (2002).
24. Kim, S. *et al.* PTEN and TNF-alpha regulation of the intestinal-specific Cdx-2 homeobox gene through a PI3K, PKB/Akt, and NF-kappaB-dependent pathway. *Gastroenterology* **123**, 1163–1178 (2002).
25. Dreyling, M.H. *et al.* Alterations of the cyclin D1/p16-pRB pathway in mantle cell lymphoma. *Cancer Res.* **57**, 4608–4614 (1997).
26. Xiu, M. *et al.* The transcriptional repressor HBP1 is a target of the p38 mitogen-activated protein kinase pathway in cell cycle regulation. *Mol. Cell. Biol.* **23**, 8890–8901 (2003).
27. McPherson, J. *et al.* A physical map of the human genome. *Nature* **409**, 934–941 (2001).
28. Watson, S.K. *et al.* Methods for high throughput validation of amplified fragment pools of BAC DNA for constructing high resolution CGH arrays. *BMC Genomics* **5**, 6 (2004).