

Short Communication

High-resolution array CGH increases heterogeneity tolerance in the analysis of clinical samples

Cathie Garnis^{a,*}, Bradley P. Coe^{a,1}, Stephen L. Lam^b, Calum MacAulay^b, Wan L. Lam^a

^aDepartment of Cancer Genetics, Canada

^bDepartment of Cancer Imaging, British Columbia Cancer Research Centre, 601 West 10th Avenue, Vancouver, Canada BC V5Z 1L3

Received 19 September 2004; accepted 28 February 2005

Available online 13 April 2005

Abstract

Recent advances in array comparative genomic hybridization (array CGH) technology are revolutionizing our understanding of tumor genomes. Marker-based arrays enable rapid survey at megabase intervals, while tiling path arrays examine the entire genome in unprecedented detail. Tumor biopsies are typically small and contain infiltrating stromal cells, requiring tedious microdissection. Tissue heterogeneity is a major barrier to high-throughput profiling of tumor genomes and is also an important consideration for the introduction of array CGH to clinical settings. We propose that increasing array resolution will enhance detection sensitivity in mixed tissues and as a result significantly reduce microdissection requirements. In this study, we first simulated normal cell contamination to determine the heterogeneity tolerance of array CGH and then validated this detection sensitivity model on cancer specimens using the newly developed submegabase resolution tiling-set (SMRT) array, which spans the human genome with 32,433 overlapping BAC clones.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Array CGH; Normal cell contamination; Microdissection; SMRT array; Tiling path array; Matrix CGH; Tissue heterogeneity

Array comparative genomic hybridization (array CGH) detects segmental DNA copy number gains and losses in tumor genomes. This is achieved by the competitive hybridization of differentially labeled reference and sample genomic DNA to specific genomic loci spotted in an array format, facilitating high-resolution scanning for genetic alterations [1–6]. To adapt whole-genome array CGH for high-throughput analysis of tumor genomes, especially in a clinical setting, this technology would have to be not only applicable to formalin-fixed paraffin-embedded tissue specimens but also able to tolerate tissue heterogeneity. Tumor samples are typically highly heterogeneous, containing multiple normal cell types in addition to the cancer cells of interest. Since array CGH identifies genetic alterations by comparing DNA copy number of the cancer cells against those of normal

diploid cells, normal cell contamination in a tumor specimen would compromise detection sensitivity.

Microdissection of specific cell populations is a common approach to overcoming tissue heterogeneity. The utility of microdissected archival material in array CGH studies is well documented; however, this approach often requires genomic DNA amplification to yield sufficient material for hybridization [7–9]. The time-consuming microdissection requirement hampers the broad application of this technique and its utility as a high-throughput technology. In this report we investigated the heterogeneity tolerance of array CGH and showed that increasing array density improves detection sensitivity directly.

To determine the maximum amount of contaminating normal cells tolerable while allowing detection of single-copy amplifications and deletions, we simulated heterogeneity by mixing precise proportions of male (X) and female (XX) DNA samples (Supplemental Fig. 1A) and then validated our model in clinical specimens. The submegabase resolution tiling-set (SMRT) array was used for assaying

* Corresponding author. Fax: +1 (604) 877 6155.

E-mail address: cgarnis@bccrc.ca (C. Garnis).

¹ These authors contributed equally.

detection sensitivity [4,10]. The SMRT array consists of 32,433 bacterial artificial chromosomes arranged in a tiling path that spans the entire genome. Array hybridization protocols were performed as previously described [4]. Briefly, for the heterogeneity titration series 400 ng of test and reference DNA were separately labeled through a random priming reaction with cyanine 3 and cyanine 5, respectively. The probes were precipitated and then combined, denatured, and blocked in a solution containing 100 μ g Cot-1 DNA in 45 μ l DIG Easy hybridization solution (Roche, Laval, QC, Canada), sheared herring sperm DNA (Invitrogen, Burlington, ON, Canada), and yeast tRNA (Calbiochem, Mississauga, ON, Canada). Probe hybridization to the SMRT array occurred over 36 h at 45°C.

In our simulation experiments, first we observed the expected linear approach to a 1:1 average signal ratio for the X loci as the level of contaminating normal cells increased. Due to the increase in overlap between the ratio distributions between the X chromosome and autosome it became apparent that thresholds would not be appropriate for identifying alterations due to the large percentage of clones that would be falsely classified and that small alterations would be more difficult to detect than alterations spanning a large number of clones (Supplemental Figs. 1B–1D).

Second, to model the impact of heterogeneity on the probability of detecting an alteration of a given size, we utilized segments of the X chromosome from the contamination hybridizations to model single-copy gains and losses of varying sizes within the autosome. This model was based on using Welch's approximate *t* test to compare every 1, 2, 4, 8, 16, 32, 64, 128, 256, and 512 clone window over the entire autosome to the X chromosome and vice versa. Remarkably, we distinguished a 4 clone single-copy loss (~0.4 Mb) on the SMRT array under 50% contamination and a 64 clone deletion (~6.4 Mb) under 75% contamination (Fig. 1A). As our results are based solely on the number of clones altered and not the genomic size of an alteration we can infer that a 1-Mb-resolution CGH array compared to the 0.1-Mb-resolution SMRT CGH array would follow the same detection probabilities and be limited to detecting a 4-Mb single-copy loss under 50% contamination and 64-Mb single-copy loss under 75% contamination. It appears that this model applies to both single-copy loss and single-copy gain scenarios (Fig. 1). The more measurements performed over a sequence improved the confidence in detecting alterations, supporting the concept that increasing array resolution reduces the need for microdissection.

We verified the modeled effect of heterogeneity on detection sensitivity using a clinical lung cancer specimen. Fig. 2A shows an H&E-stained section from a tumor from a male patient. Histological assessment suggested a mix of 30% tumor cells infiltrated with 70% stromal cells and lymphocytes as well as carbon deposits. Enumeration of tumor and normal cell nuclei in the displayed area counted $28 \pm 15\%$ cancer cells, in agreement with the initial

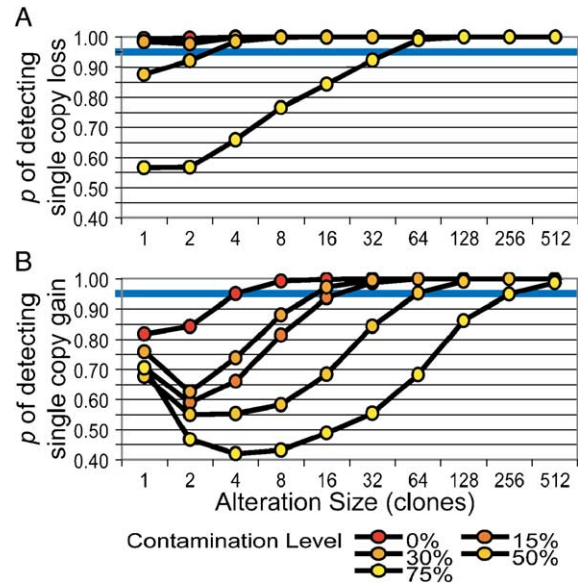


Fig. 1. Minimal detectable alteration sizes. To calculate the probability of detecting a segmental alteration of a particular size at a particular normal cell contamination level we first divided the autosome and X chromosome into segments of 1, 2, 4, 8, 16, 32, 64, 128, 256, and 512 clones (each clone representing three measurements). The autosomal segments were used to simulate areas of retention within the X chromosome and the X-chromosome segments were used to simulate alterations within the autosome. This was accomplished by first using Welch's approximate *t* test to determine if a particular segment from the autosome could be identified as distinct from the X chromosome with a *p* value of 0.05. The frequency with which the autosomal segments were not identified as being distinct from the X chromosome defined the percentage of segments incorrectly detected as altered. The inverse test of segments from the X chromosome being compared to the autosome was used to determine the percentage of segments correctly detected as altered. Due to the different variances between the X chromosome and the autosome distributions as well as the nature of the *t* test, we cannot assume that the fraction of clones correctly and incorrectly identified as altered will sum to exactly 100%. As such, we calculate the probability of detecting an alteration at a particular contamination level and alteration size as Probability of detection = Fraction correctly identified as altered/(Fraction correctly identified as altered + Fraction incorrectly identified as altered). Due to the shift from an inter- to an intracell measure of variance associated with alterations spanning only a small number of clones the *p* values reported for the smallest alterations exhibit a slight overestimate in detection probability, particularly in the highest contamination levels. This is most apparent in the single-copy gain scenario as single-copy gains (3:2 allele ratio) exhibit less ratio separation from normal than single-copy losses (1:2 allele ratio). However, since these probabilities are well below the threshold for reliable detection the overestimate does not affect our results.

histological assessment. DNA extracted from this mixed-cell population was cohybridized against differentially labeled normal female DNA (as a reference) onto the SMRT CGH array. Analysis of X-chromosome loci yielded the expected average 0.5 \log_2 ratio (Fig. 2B). Even with the compromised DNA quality from formalin-fixed paraffin-embedded archival material and the high level of normal cell contamination, copy number changes in the tumor DNA were apparent. Large alterations, such as gain of 5p and loss of 5q, as well as high-level amplifications (for example, the

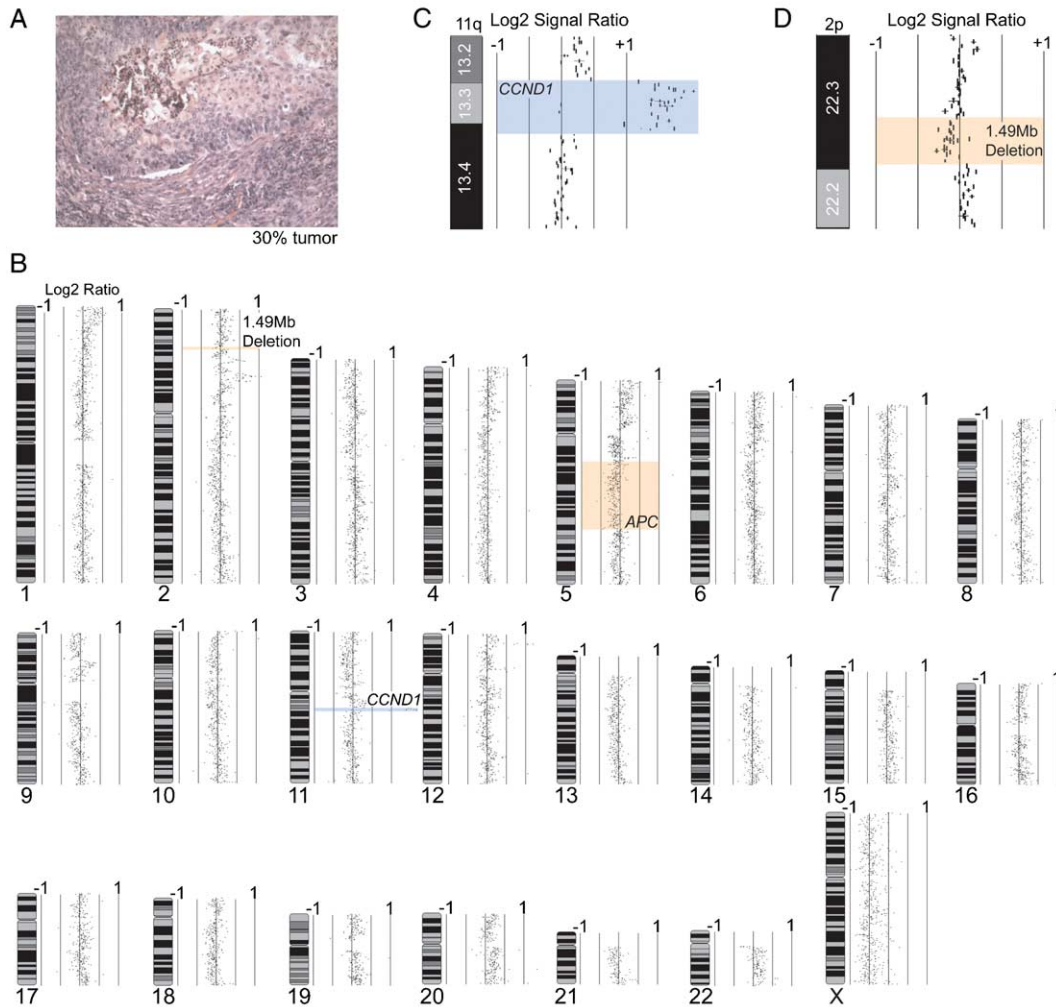


Fig. 2. SMRT array CGH profile of an archival squamous cell lung tumor containing ~30% tumor cells. (A) H&E-stained squamous cell lung carcinoma tissue section containing ~30% tumor cells by histological evaluation. (B) SeeGH karyogram [11] of the DNA extract from the tissue in (A). Hybridization was performed as described by Ishkanian et al. except we used 100 ng of tumor DNA against 100 ng of female reference DNA [4]. Each black dot/line represents a single BAC clone marking its genomic position and \log_2 signal ratio value. Regions of interest are highlighted. (C) Magnified SeeGH view of the highly amplified region on chromosome 11q containing *CCND1*. (D) Magnified SeeGH view of a 1.49-Mb single copy number deletion on chromosome 2p.

2.5-Mb *CCND1* region) were readily detected (Figs. 2B and 2C). Presumably multicopy amplifications are easily detectable regardless of clone density.

Given that only ~30% of the cells in our sample are cancerous, the signal ratio for a single-copy loss within these cells is expected to be 1.7:2 (0.23 \log_2 ratio). The signal ratio associated with the 5q loss, which contains the well-studied *APC* gene, fits this expectation. More interestingly, we also observed a small 1.49-Mb single-copy deletion (across 22 clones) at 2p22 with the same ratio (Fig. 2D). Significantly, this observation fits well into our model of detection sensitivity in mixed tissue, which predicted the ability to identify a loss spanning between 4 clones (the threshold for 50% contamination) and the predicted 64-clone threshold for 75% contamination (Fig. 1A). With the ability to detect copy number changes with only 30% tumor cell content, tedious microdissection is not required when using the SMRT array. This is crucial for the

development of array CGH as a clinical screening tool for use in a high-throughput setting.

In conclusion, we have determined the sensitivity of array CGH to single-copy changes under heterogeneous conditions. Our results show that increasing array resolution improves directly detection sensitivity to smaller alterations in heterogeneous tissue. Clinical tumor samples are typically highly heterogeneous, containing mixed cell types in addition to the malignant cells. As a result microdissection is often required before genome-wide analysis with current techniques, posing a barrier to high-throughput profiling of tumor genomes and the introduction of array CGH to diagnostic cancer cytogenetics. The extremely high resolution of the SMRT array, and hence its sensitivity to small alterations under high levels of contamination, greatly reduces the need for time-consuming microdissection, removing a major hurdle to the introduction of array CGH into the clinical setting.

Acknowledgments

We thank Dr. Jean LeRiche for evaluation of lung specimens and Dr. Carolyn Brown for X chromosome discussions. This work was supported by grants from the National Cancer Institute of Canada, Genome Canada/British Columbia, Canadian Institute of Health Research, National Institute of Dental and Craniofacial Research (RO1DE015965 and RO1DE13124), National Science and Engineering Research Council (C.G.), and Michael Smith Foundation for Health Research scholarships to C.G. and B.P.C.

Appendix A. Supplementary data

Supplementary data for this article may be found on ScienceDirect.

References

- [1] J.R. Pollack, C.M. Perou, A.A. Alizadeh, M.B. Eisen, A. Pergamenschikov, C.F. Williams, et al., Genome-wide analysis of DNA copy-number changes using cDNA microarrays, *Nat. Genet.* 23 (1999) 41–46.
- [2] A.M. Snijders, N. Nowak, R. Segraves, S. Blackwood, N. Brown, J. Conroy, et al., Assembly of microarrays for genome-wide measurement of DNA copy number, *Nat. Genet.* 29 (2001) 263–264.
- [3] R. Lucito, J. Healy, J. Alexander, A. Reiner, D. Esposito, M. Chi, et al., Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation, *Genome Res.* 13 (2003) 2291–2305.
- [4] A.S. Ishkanian, C.A. Malloff, S.K. Watson, R.J. De Leeuw, B. Chi, B.P. Coe, et al., A tiling resolution DNA microarray with complete coverage of the human genome, *Nat. Genet.* 36 (2004) 299–303.
- [5] C. Garnis, C. Baldwin, L. Zhang, M.P. Rosin, W.L. Lam, Use of complete coverage array comparative genomic hybridization to define copy number alterations on chromosome 3p in oral squamous cell carcinomas, *Cancer Res.* 63 (2003) 8582–8585.
- [6] B. Carvalho, E. Ouwerkerk, G.A. Meijer, B. Ylstra, High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides, *J. Clin. Pathol.* 57 (2004) 644–646.
- [7] Y. Daigo, S.F. Chin, K.L. Gorringer, L.G. Bobrow, B.A. Ponder, P.D. Pharoah, et al., Degenerate oligonucleotide primed-polymerase chain reaction-based array comparative genomic hybridization for extensive amplicon profiling of breast cancers: a new approach for the molecular analysis of paraffin-embedded cancer tissue, *Am. J. Pathol.* 158 (2001) 1623–1631.
- [8] P.P. Massion, W.L. Kuo, D. Stokoe, A.B. Olshen, P.A. Treseler, K. Chin, et al., 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 (2002) 3636–3640.
- [9] G. Wang, C. Brennan, M. Rook, J.L. Wolfe, C. Leo, L. Chin, et al., Balanced-PCR amplification allows unbiased identification of genomic copy changes in minute cell and tissue samples, *Nucleic Acids Res.* 32 (2004) e76.
- [10] R.J. De Leeuw, J.J. Davies, A. Rosenwald, G. Bebb, R.D. Gascoyne, M.J. Dyer, et al., Comprehensive whole genome array CGH profiling of mantle cell lymphoma model genomes, *Hum. Mol. Genet.* 13 (2004) 1827–1837.
- [11] B. Chi, R.J. DeLeeuw, B.P. Coe, C. MacAulay, W.L. Lam, SeeGH—A software tool for visualization of whole genome array comparative genomic hybridization data, *BMC Bioinform.* 5 (2004) 13.