Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 85 (2005) 790–793

GENOMICS

www.elsevier.com/locate/ygeno

Short Communication

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

Cathie Garnis[a,*,1], Bradley P. Coe[a,1], Stephen L. Lam[b], Calum MacAulay[b], Wan L. Lam[a]

[a]*Department of Cancer Genetics, Canada*

[b]*Department 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.

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.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.02.015

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 μg Cot-1 DNA in 45 μl 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 ± 15% cancer cells, in agreement with the initial

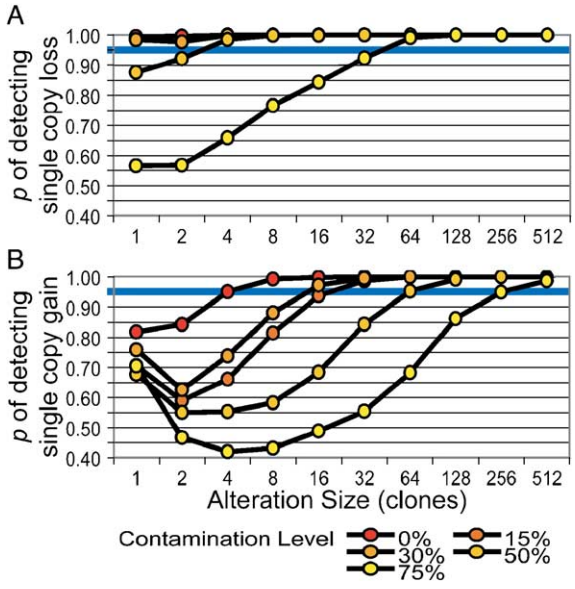

Fig. 1. Minimal detectable alteration sizes. To calculate the probability of detecting a segmental alteration of a particular size at a particular normal cell contamination level we first divided the autosome and X chromosome into segments of 1, 2, 4, 8, 16, 32, 64, 128, 256, and 512 clones (each clone 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 intraclone 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