Genomic Analysis of Tumors by Array Comparative Genomic Hybridization: More Is Better

To the Editor:

We read with interest the article by Baldwin et al. (1), which describes the use of an array of 32,433 bacterial artificial chromosome clones tiled across the genome (2, 3) to map genomic aberrations in 20 oral squamous cell carcinoma samples by array comparative genomic hybridization (array CGH). The authors conclude that improved detection sensitivity and resolution of tiling arrays is necessary for analysis of tumor genomes. No one would argue against the benefits of higher-resolution copy number measurements. Indeed, regions of aberration have been narrowed using arrays, providing high-density regional coverage (4, 5), and direct comparison shows that tiling arrays provide more information on aberrations that oftentimes are missed or detected by only single clones on lower-resolution arrays. Nevertheless, comparison with the study of Snijders et al. (6), which used bacterial artificial chromosome arrays with megabase resolution to analyze genomic aberrations in 89 oral squamous cell carcinoma samples, shows that another important contributor to the biological utility of an experiment is the inclusion of sufficient numbers of samples distributed over the breadth of tumor subtypes.

Several examples indicate the interplay of array resolution and sample cohort in affecting biological conclusions. Snijders et al. (6) showed that oral squamous cell carcinomas differ widely in frequency of copy number aberrations, with hierarchical clustering based on genomic aberrations indicating at least two subtypes. Thus, the distribution of patients among these groups will affect overall estimates of aberration frequencies and is more likely to explain the difference in frequency of *EGFR* amplification reported by the two groups. Differences in the classification of loci as being "amplified" may also have contributed, but the statement by Baldwin et al. (1) that the higher frequency of *EGFR* amplification in their data set is evidence of increased detection sensitivity is unlikely to be the cause, because the array used by Snijders et al. contained a specific clone for the *EGFR* locus. Thus, none of these amplifications could be missed by the lower-resolution array. [Baldwin et al. also incorrectly cite the number of such amplifications found by Snijders et al. (6) as 4 of 89 instead of 10 of 89.] In addition, in at least one case, the analysis of more tumors by Snijders et al. (6) provided greater biologically relevant genomic resolution than Baldwin et al., narrowing the minimal amplified region at 11q22.3 to 0.8 Mb (excluding the matrix metalloproteinases distal to *MMP7*) compared with the 1.24 Mb interval reported by Baldwin et al.(1). On the other hand, the tiling array provided greater refinement of the *CDK6* amplicon.

We note that regardless of resolution, it may be necessary to validate some results by other means, particularly those involving single clones. For example, Baldwin et al. (1) reported that the gain of a single clone RP11-338N5 at 7p12.3-p13 (*TENS1*) is present in 60% of cases, potentially indicating the significance of this gene in oral cancer. Nevertheless, whereas Baldwin et al. (1) map RP11-338N5 to 7p, the National Center for Biotechnology Information (NCBI) database places it on 3p, and we have it located at 8q24.3 on our tiling array (3). In our experience, ratio changes reported by this clone agree with location at 8q24.3. This position would also be consistent with the high frequency of gain (60%) reported by Baldwin et al. (1) because +8q is common in oral cancer.

Clearly, measuring the largest possible number of tumors on the highest-resolution arrays will provide the greatest information. Given the range of aberration sizes in cancer, properly annotated primary data of any resolution are valuable if readily accessible and will allow meta-analyses using consistent analytic criteria to be done. Similarly, a number of groups are carrying out array CGH using the set of tiling array clones assembled by Kryzwinski et al. and access to data will facilitate community annotation of copy number variants, as well as updating mapping information. Snijders et al. published their data; unfortunately, Baldwin et al. did not. We urge the editors of *Cancer Research* to make database submission (e.g., ArrayExpress, NCBI GEO) of primary data and associated patient characteristics a requirement for publication of genomic data.

Donna G. Albertson
Antoine M. Snijders
Jane Fridlyand
Richard Jordan
Daniel Pinkel
Brian L. Schmidt
Comprehensive Cancer Center, University of California San Francisco, San Francisco, California

References

1. Baldwin C, Garnis C, Zhang L, Rosin MP, Lam WL. Multiple microalterations detected at high frequency in oral cancer. Cancer Res 2005;65:7561–7.
2. 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.
3. Krzywinski M, Bosdet I, Smailus D, et al. A set of BAC clones spanning the human genome. Nucleic Acids Res 2004;32:3651–60.
4. Albertson DG, Ylstra B, Segraves R, et al. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat Genet 2000;25:144–6.
5. 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;76:750–2.
6. Snijders AM, Schmidt BL, Fridlyand J, et al. Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. Oncogene 2005;24:4232–42.

In Response:

As array CGH grows to be a valued genetic tool, it becomes increasingly important to understand the factors that affect biological interpretation of array CGH data. Albertson et al. raised the issue of the interplay of array resolution and sample cohort in affecting biological conclusions, citing good examples of this interplay.

Increased array resolution not only improves precision of defining the alteration boundaries but also provides confidence and sensitivity in detecting alterations as multiple array elements rather than single clones are affected by the alteration. For example, the

doi:10.1158/0008-5472.CAN-05-3611

TENS1 locus at 7p12.3-p13 described by Baldwin et al. (1) represents a region defined by overlapping clones minimally encompassing RP11-338N5. Likewise, a large sample cohort size will facilitate fine mapping and potentially the delineation of oral cancer subtypes. The suggestion of Albertson et al. that there may be different subtypes of oral cancer with different levels of *EGFR* amplification is intriguing. Whether or not this explains the frequency differences observed by Snijder et al. (10/89; ref. 2) and by Baldwin et al. (8/20; ref. 1) remains to be determined. It is important to note that there are other biological factors that could mask detection of genetic alterations, especially low copy number events. A prime example is tissue heterogeneity, which is characteristic of tumors. Compromised detection sensitivity of focal alterations may have great effect on megabase interval arrays as multiple measurements would span large genomic distances (3). Whereas tissue microdissection can enrich for cells with malignant appearance, events present in subpopulations may be missed. In addition, increased cohort size improves the probability of detecting recurrent segmental copy number alteration, as other mechanisms such as epigenetic changes, recombination genetic events, and mutations at the sequence level can also contribute to the frequency of deregulation of specific genes.

As more array CGH data become available, the possibility of cross-study meta-analysis of results generated by multiple groups on the same tumor type becomes feasible. One barrier to such analysis is the technical variation in array platforms, with different formats, detection sensitivities, and ambiguous definition of resolution, complicating cross-platform comparisons. Another barrier is the lack of consensus on array CGH data deposition. Currently, the requirements for deposition of genomic array data vary among scientific journals. Standards for deposition in public databases were designed for expression profiles and are often inappropriate for reporting segmental copy number status. More significantly, whole-genome array CGH data typically consist of tens of thousands of spots, usually in replicates; raw data deposited into databases such as NCBI GEO may not be readily useful to most cancer researchers as they cannot view the data in a meaningful way, let alone interpret the data and use the information. Currently, only investigators with expertise in array CGH can readily take advantage of the wealth of information from the deposited data.

For "more to be better," we need not only to increase resolution and sample size but also to develop a worldwide consensus in format for data deposition and easy-to-use software for visualization and analysis of data.

Corisande Baldwin
Cathie Garnis
Lewei Zhang
Miriam Rosin
Wan L. Lam
British Columbia Cancer Research Centre
Vancouver, British Columbia, Canada

References

1. Baldwin C, Garnis C, Zhang L, Rosin MP, Lam WL. Multiple microalterations detected at high frequency in oral cancer. Cancer Res 2005;65:7561–7.
2. Snijders AM, Schmidt BL, Fridlyand J, et al. Rare amplicons implicate frequent deregulation of cell fate specification pathways in oral squamous cell carcinoma. Oncogene 2005;24:4232–42.
3. 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;85:790–3.

Genomic Analysis of Tumors by Array Comparative Genomic Hybridization: More Is Better

Donna G. Albertson, Antoine M. Snijders, Jane Fridlyand, et al.

Cancer Res 2006;66:3955-3956.