Memorial Sloan-Kettering Cancer Center
Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology & Biostatistics Working Paper Series

Year 2007 *Paper* 13

Statistical Evaluation of Evidence for Clonal Allelic Alterations in array-CGH Experiments

Colin B. Begg* Kevin Eng†
Adam Olshen‡ E S. Venkatraman**

*Memorial Sloan-Kettering Cancer Center, beggc@mskcc.org

†Memorial Sloan-Kettering Cancer Center, engk@mskcc.org

‡Memorial Sloan-Kettering Cancer Center, Dept. of Epidemiology and Biostatistics, olshena@biostat.ucsf.edu

**Memorial Sloan-Kettering Cancer Center, venkatre@mskcc.org

http://biostats.bepress.com/mskccbiostat/paper13

Statistical Evaluation of Evidence for Clonal Allelic Alterations in array-CGH Experiments

Colin B. Begg, Kevin Eng, Adam Olshen, and E S. Venkatraman

Abstract

In recent years numerous investigators have conducted genetic studies of pairs of tumor specimens from the same patient to determine whether the tumors share a clonal origin. These studies have the potential to be of considerable clinical significance, especially in clinical settings where the distinction of a new primary cancer and metastatic spread of a previous cancer would lead to radically different indications for treatment. Studies of clonality have typically involved comparison of the patterns of somatic mutations in the tumors at candidate genetic loci to see if the patterns are sufficiently similar to indicate a clonal origin. More recently, some investigators have explored the use of array CGH for this purpose. Standard clustering approaches have been used to analyze the data, but these existing statistical methods are not suited to this problem due to the paired nature of the data, and the fact that there exists no "gold standard" diagnosis to provide a definitive determination of which pairs are clonal and which pairs are of independent origin. In this article we propose a new statistical method that focuses on the individual allelic gains or losses that have been identified in both tumors, and a statistical test is developed that assesses the degree of matching of the locations of the markers that indicate the endpoints of the allelic change. The validity and statistical power of the test is evaluated, and it is shown to be a promising approach for establishing clonality in tumor samples.

December 11, 2006

Statistical Evaluation of Evidence for Clonal Allelic Alterations in array-CGH Experiments

Colin B. Begg*, Kevin H. Eng, Adam B. Olshen, E.S. Venkatraman

Department of Epidemiology and Biostatistics
Memorial Sloan-Kettering Cancer Center
307 East 63rd Street
New York, NY 10021
*Corresponding Author: beggc@mskcc.org, 646-735-8108, 646-735-0009 (FAX)

SUMMARY

In recent years numerous investigators have conducted genetic studies of pairs of tumor specimens from the same patient to determine whether the tumors share a clonal origin. These studies have the potential to be of considerable clinical significance, especially in clinical settings where the distinction of a new primary cancer and metastatic spread of a previous cancer would lead to radically different indications for treatment. Studies of clonality have typically involved comparison of the patterns of somatic mutations in the tumors at candidate genetic loci to see if the patterns are sufficiently similar to indicate a clonal origin. More recently, some investigators have explored the use of array CGH for this purpose. Standard clustering approaches have been used to analyze the data, but these existing statistical methods are not suited to this problem due to the paired nature of the data, and the fact that there exists no "gold standard" diagnosis to provide a definitive determination of which pairs are clonal and which pairs are of independent origin. In this article we propose a new statistical method that focuses on the individual allelic gains or losses that have been identified in both tumors, and a statistical test is developed that assesses the degree of matching of the locations of the markers that indicate the endpoints of the allelic change. The validity and statistical power of the test is evaluated, and it is shown to be a promising approach for establishing clonality in tumor samples.

Keywords: clonality; array CGH; permutation test.

1. INTRODUCTION

Clinical investigators frequently conduct experiments to determine whether tumors share a clonal origin. That is, one wishes to determine if the tumors are derived from the same "clonal" cell. In addition to informing the interpretation of experiments about the mechanistic development of cancers, comparisons of samples of cells from pairs of tumors from the perspective of clonality can have important clinical implications. For example, a patient treated effectively for a localized primary head and neck cancer may at a later date present with a solitary lung nodule. If the nodule is a localized second primary lung cancer it can be treated effectively by surgery, though lung surgery is risky and very invasive. On the other hand, if the tumor is a metastasis from the head/neck primary, the prognosis of the patient is necessarily poor, as the cancer will almost certainly have also metastasized to other parts of the body (even though these other metastases may not yet be detectable). In this case invasive surgery would impose needless risks and morbidity on a patient who will have relatively little time left to live.

Currently, pathologists make this call on the basis of histopathologic characteristics, but this is not a definitive strategy. Recently, many investigators have begun to study the issue by using molecular profiling of the tumors. For example, investigators studying lung cancer have used microsatellite markers to distinguish microsatellite instability (Huang 2001, Dacic 2005, Geurts 2005, Leong 1998, Shin 2001) and several investigators have also used mutational analysis of the important cancer genes p53 and/or K-ras (Hiroshima 1998, Holst 1998, Lau 1997, Shimizu 2000, Shin 2001, Murase 2003, Matsuzoe 1999, Sozzi 1995, van Rens 2002). Similar studies have been conducted to distinguish contralateral breast cancers fron metastases, and in other cancer sites (Imyanitov 2002, Regitnig 2004, Kollias 2000, Janschek 2001, Tse 2003, Schlechter 2004, Stenmark-Askmalm 2001, Chunder 2004). These studies have evaluated clonality in a range of clinical settings, including the comparison of synchronous or metachronous multiple primaries, comparisons of primaries with

metastatic tumors, and studies of multiple potential clones within tumors that harbor multiple histologies. By studying the mutational pattern, one can establish a genetic fingerprint of the tumor. When the mutational profiles of two apparently independent primary tumors from the same patient are compared, it is possible in principle to see whether these genetic fingerprints are sufficiently similar that we can determine with confidence that they share a clonal origin, i.e. the second primary is really a metastasis from the first primary.

The comparison of mutational profiles of tumors to determine clonality is a challenging statistical problem, and a number of authors have proposed techniques for this purpose. The fundamental goal is to examine the profiles of the two tumors to see whether the evidence favors a clonal versus an independent origin for the tumors. In earlier work we examined candidate statistical tests for this purpose, based on the setting in which the frequencies of mutational events (usually LOH) at candidate loci are assessed for correlation, with a view to determining if the correlation exceeds the level that is plausible on the basis of chance (Begg et al. 2006). These tests take advantage both of the information in the correlation of mutational events and the extent to which common mutations at the same locus occur on the same parental allele. The tests have been shown to be reasonably powerful provided that information is available from a considerable number of candidate genetic loci that experience mutational events with reasonably high frequency in the cancer under study, and that the signal is relatively strong, i.e. the preponderance of the mutations occur in the clonal phase of development. Other authors have approached this problem in different ways. For example Sieben et al. (2003) and Brinkmann et al. (2004) both construct likelihood ratios to distinguish the evidence favoring the two hypotheses, though the construction is somewhat different in each case. A different approach was advocated in earlier work by Kuukasjarvi et al. (1997), who proposed a measure of clonal relatedness based on the frequency of occurrence of concordant mutations in the tumors, and this measure has been used by other authors such as Jiang et al. (2005) and Goldstein et al. (2005a,b).

The preceding methods are all based on the setting in which we observe mutations in a pre-specified set of candidate markers in each tumor, and we evaluate the collective concordance of these markers. Since the common somatic mutations in tumors are frequently losses or gains of segments of DNA, the issue of clonality can be studied for the entire genome using array technology, specifically array comparative genomic hybridization (ACGH) (Pinkel et al. 1998). By scanning the entire genome for copy number changes this technology has the potential to provide a more comprehensive comparison of the two tumors, and to provide insights beyond those available from studies of LOH at individual markers. In particular, ACGH can detect both copy number gains and losses, and it can pinpoint the places in the genome where these gains and losses begin and end, although it does not distinguish the specific allele on which the loss or gain occurs. Despite the potential precision with which a specific allelic gain or loss is determined, statistical methods used in this context have employed strategies that simply count mutational events, as in the methods described above for studies based on candidate loci. For example, investigators have used the data from the arrays to define the presence or absence of, say, LOH at the level of the chromosome arm (Jiang et al. 2005) or chromosome band (Teixeira et al. 2004) in order to define the unit of analysis for the use of statistical tests or clustering algorithms, or for the computation of the clonal relatedness index.

Our strategy in this article is to take advantage of the aspect of ACGH data that distinguishes the nature of the information from that obtained in studies involving candidate genetic loci, namely the granularity of the information regarding the allelic gains and losses. This feature of the data provides the ability to pinpoint the start and stop regions of these allelic changes, with a view to determining an exact match between the mutations on the two tumors. In so doing, our method offers the potential to establish clonality on the basis of a single allelic gain or loss simply on the basis of the closeness of the match, regardless of the possible absence of concordance of bystander somatic events that may have occurred after the two tumors developed as separate clones. Even though a clonal allelic gain or loss is necessarily identical on both tumors, noise in the measurement of the CGH markers results in statistical uncertainty regarding the precise

start and stop points of the allelic change. The goal of our approach is to see if the estimated start and stop points of allelic events on the two tumors that appear to be possibly clonal are sufficiently close that events of such a degree of similarity are unlikely to have occurred by chance. The construction of a valid statistical test for this purpose is challenging for several reasons, including the fact that gains or losses of independent origin are likely to be correlated, since their locations on the genome do not occur independently.

2. METHODS

Array comparative genomic hybridization is a technique for determining allelic copy number changes across the entire genome. Samples of tumor tissue and normal tissue are differentially labeled and co-hybridized to a slide containing genomic markers with a view to determining regions of allelic loss or gain in the tumor. Several array techniques have been developed to increase resolution in order to permit identification of relatively small allelic changes (Pinkel and Albertson 2005). The result is a linear map of marker values which represents the copy number in the tumor at sequential locations across the genome.[The measurements are relative but they reflect the absolute copy number of the tumor.] These marker values are, of course, subject to random variation, and so the determination of regions of gain or loss requires statistical analysis, and several techniques for this purpose have been developed (reviewed in Lai et al. 2005).

Our goal in this article is to compare the ACGH profiles from two tumors in the same patient, with a view to determining whether the tumors arose independently, or were derived from a single clonal cell that experienced one or more of the observed allelic changes. If there are "clonal" mutations in the two tumors then these are necessarily identical. Thus the estimated regions of gain or loss must be very similar in the two tumors, though estimates of the precise markers at which the regions of gain or loss begin and end will not necessarily be identical, due to random fluctuations in the marker values, and the resulting imprecision of the methods in identifying the regions of gain or loss.

An example of data from two tumors from the same patient is provided in Figure 1. The horizontal lines show estimated regions of constant copy number, and discontinuities within chromosomal arms indicate positions of presumed allelic loss or gain. The key question is whether one (or more) of the somatic events can be convincingly demonstrated to be clonal, on the basis of the exactness of the match. The arrows in the diagram point to changes that look plausibly similar at first glance, and our test is designed to examine these plausible matches individually. In other words we reduce the problem to an evaluation of the likelihood of a match for the specific concordant allelic changes observed in both tumors. Our method addresses whether any of these changes are close enough to convince us that the tumor is of clonal origin.

Regions of gain or loss are evaluated within chromosome arms, and these frequently involve the gain or loss of an allele across the entire chromosome arm. Other patterns include partial arm alterations, simple alterations and complex alterations. A simple alteration is one in which a single allelic gain or loss occurs within the chromosome arm. Partial arm alterations are, in essence, simple alterations in which one of the two endpoints occurs at a boundary. A complex pattern of alterations involves more than one gain or loss (see Figure 2). As background for this article we examined high-resolution ACGH data from a series of 38 patients with diffuse large B-cell lymphoma diagnosed between 1984 and 1998 (Chen et al. 2006). Among all of the chromosomal arms examined, changes were identified in 18%. Excluding whole arm changes, the preponderance of alterations were either partial arm (46%) or simple (43%) changes. Since the presence of a whole arm alteration is relatively common it is not improbable that the same alteration will occur by chance on both tumors. Consequently, the occurrence of a concordant whole arm gain or loss on the same chromosomal arm on both tumors will not provide strong enough evidence on its own to establish a clonal origin for the tumors, though it will be relevant in aggregating the evidence from the entire genome. Most of the remaining types of mutational events are either partial arm alterations or simple alterations. We focus in this article on the challenge of comparing concordant simple or partial alterations that occur on the same chromosome arm of the

two tumors. By "concordant", we mean that either both changes are gains or both are losses, and that each one represents a statistically significant allelic change on the basis a statistical test for detecting allelic change. We develop a new statistical test for comparing the estimated start and stop markers in the two tumors. We regard the resulting test as being the building block for interrogating the entire pattern of mutations on the two tumors.

Our general strategy is conceptually straightforward. After defining notation in the next section, we propose a test statistic that represents the "closeness" of the observed allelic changes on the two tumors. We then outline an approach to determining a reference distribution for this statistic under the null hypothesis that the two mutations occurred independently. This is a challenging task for several reasons that are discussed in detail. Later we examine the validity and power of the test using simulations, and apply it to data from various patients.

2.1 Notation and Test Statistic

Let x_{uk} represent the measurement of the u^{th} marker of the k^{th} tumor, where $u = 1,..,n,$ and $k = 1,2.$ Consider the setting in which there is a concordant gain (or loss) on the two tumors. Let the copy number change begin at marker i_k and end at marker j_k for the k^{th} tumor. That is, markers i_k through j_k, inclusive, represent the markers of allelic gain (or loss). If the tumors are clonal then $i_1 = i_2$ and $j_1 = j_2,$ and this represents the alternative (clonal) hypothesis in our formulation. The null hypothesis is that the regions of gain or loss have arisen independently and consequently that the endpoints of the allelic change are very unlikely to be identical. [Note: we recognize that an exact match could occur by chance, though this is highly improbable at the levels of granularity of the arrays in which we are interested.]

The endpoints need to be estimated using an algorithm for detecting regions of allelic change. In our simulations we use the CBS (circular binary segmentation)

algorithm (Olshen et al., 2004), although in principle any similar method could be used. The CBS algorithm obtains estimates of the endpoints, denoted $\hat{i}_k$ and $\hat{j}_k$, by maximizing the mean difference between the markers in the hypothesized region of gain or loss and the remaining markers. In fact, the CBS algorithm chooses $\hat{i}_k$ and $\hat{j}_k$ to maximize $|z_{ijk}|$ where

$$z_{ijk} = \left(\frac{1}{j-i+1} + \frac{1}{n-j+i-1} \right)^{-1/2} \left(\frac{S_{jk} - S_{i-1,k}}{j-i+1} - \frac{S_{nk} - S_{jk} + S_{i-1,k}}{n-j+i-1} \right),$$

and where $S_{lk} = \sum_{u=1}^{l} x_{uk}$. [Note that in the Olshen et al. (2004) manscript the region of change is defined as $i_k + 1$ through j_k, rather than i_k through j_k.] A permutation algorithm is used to determine whether the detected change is statistically significant. In the published version of the CBS algorithm (Olshen et al. 2004) this process continues until all statistically significant change points are detected. In the strategy for this article we limit the CBS algorithm to the initial step for detecting the single, most highly significant allelic change. In our testing strategy for clonality, we start with the assumption that the CBS algorithm has detected concordant allelic changes on each tumor, both of which are statistically significant. The changes are concordant if $\delta_1 = \delta_2$, where $\delta_k = sign(\max[z_{ijk}])$. Based on these conditions our test statistic is defined as

$$t = \left| \hat{i}_1 - \hat{i}_2 \right| + \left| \hat{j}_1 - \hat{j}_2 \right|. \qquad (2.1)$$

Thus small values of t are indicative of a possible clonal mutation.

2.2 Reference Distribution

Since our test is based on the condition that two concordant allelic changes are present, our reference distribution under the null hypothesis of independence must also be based on this assumption, i.e., we assume that there exists allelic changes on both tumors that have arisen independently, and that either both represent gains or both represent losses. The reference distribution for t should then reflect the distribution of t when

(concordant) independent allelic gains or losses are generated on each tumor on the same chromosomal arm. The construction of the reference distribution is a challenging task for the following reasons.

The fact that the two sets of data being compared are required to have statistically significant observed allelic changes makes a direct permutation of the data invalid since the distribution of the markers is non-exchangeable (the gained or lost region has a different mean than the region of normal copy number). A plausible assumption of exchangeability of the marker values can be induced by subtracting the true means from the observed data. That is, the reference distribution can be generated by permuting the residuals. However, this requires knowledge of the location of the changes and the corresponding means. In our approach we use the estimated locations and means instead. Specifically, let θ_k denote the mean value for markers representing normal copy number for the k^{th} tumor, and let μ_k denote the mean marker value in the region of allelic change. We estimate these using their sample means as follows,

$$\hat{\mu}_k = \frac{1}{\hat{j}_k - \hat{i}_k + 1} \sum_{u=\hat{i}_k}^{\hat{j}_k} x_{uk}$$

$$\hat{\theta}_k = \frac{1}{n - \hat{j}_k + \hat{i}_k - 1} \left[\sum_{u=1}^{\hat{i}_k - 1} x_{uk} + \sum_{u=\hat{j}_k + 1}^{n} x_{uk} \right].$$

These are used to obtain residuals for each of the marker values:-

$$r_{uk} = x_{uk} - \hat{\theta}_k \text{ for } u < \hat{i}_k \text{ or } u > \hat{j}_k$$
$$r_{uk} = x_{uk} - \hat{\mu}_k \text{ for } \hat{i}_k < u < \hat{j}_k .$$

In order to obtain resampled data corresponding to two tumors with concordant allelic changes, the permuted residuals are added back to new mean functions that are regenerated for each permutation.

Generation of these mean functions is complicated by the fact that allelic changes will not occur at random in tumors. While chromosomal breakpoints may occur randomly in cells, the alteration is more likely to be retained if it contains a gene or genes

for which there is an advantage to having an abnormal number of copies, such as an oncogene or a tumor suppressor gene. This results in selection and clonal expansion of cells harboring specific genomic alterations. To address this phenomenon we first generate a location for a hypothetical mutational hotspot, which we presume to be located where the observed regions of allelic loss or gain on the two tumors overlap. We randomly generate new regions of allelic change for the two tumors, restricted to the set of changes that overlap the hotspot. We use the aggregate data from the two tumors to estimate the mean values for the normal and allelic change markers to increase the stability of the process.

Since the samples being compared have concordant allelic changes that have been detected as statistically significant individually, the reference distribution should also be based on samples with statistically significant changes. Hence, we use the same algorithm on the generated data for each tumor as the one used for the original data (CBS in our case) to estimate the start and stop points for the allelic changes, and to test their statistical significance. If both are significant at the same α level as the original sample and the changes are concordant, then the data set is considered to be "admissible", and the estimated endpoints are used to calculate the reference test statistic. This process is then repeated a large number of times to establish the reference distribution for the statistic t in (2.1).

The operating characteristics of this procedure depend on the signal to noise ratio as well as the properties of the procedure used for the estimation of the locations of the change-points and the means of the marker values. The ability to detect a change is a function of the signal-to-noise ratio. A lower signal-to-noise ratio results in a larger proportion of false positives among the detected changes. It also affects the test for clonality by increasing the frequency of inadmissible permutations. Also, the CBS algorithm chooses the maximum value of the CBS test statistic from among all of the possible combinations of start and stop markers for the allelic change. Since for any two pairs of start and stop markers the correlation between these statistics increases with the number of overlapping markers, the CBS algorithm tends to select smaller intervals in

preference to larger intervals when there is no real allelic change, i.e. when there is a false positive change. This preference for shorter intervals is high when the signal to noise ratio is relatively low since the preponderance of statistically significant intervals are false positives. It poses a further problem for choosing the location of the mutational hotspot since the intervals of allelic change on the two tumors are less likely to overlap.

The following algorithm for generating the reference distribution is constructed in recognition of these issues. An asterisk denotes terms representing the reference distribution:-

(1) Generate the location of the mutational hotspot h^*, where h^* is selected uniformly from the common interval, i.e. the interval between $\max(\hat{i}_1, \hat{i}_2)$ and $\min(\hat{j}_1, \hat{j}_2)$. [For simplicity we assume throughout that the hotspot occurs at a marker value, and define $\mathrm{U}(i, j)$ to represent uniform sampling of the markers between i and j, inclusive.] If the intervals do not overlap then h^* should be chosen randomly from the interval between the estimated intervals, i.e. if the second interval is higher than the first, then choose h^* randomly between $\hat{j}_1$ and $\hat{i}_2$, and vice versa. [The goal here is to accommodate two very short intervals that are close to each other but do not overlap.]

(2) Generate the "true" endpoints of the allelic changes in the reference data set: i_1^* and i_2^* sampled from $\mathrm{U}(1, h^*)$ and j_1^* and j_2^* sampled from $\mathrm{U}(h^*, n)$.

(3) Obtain $\{r_{uk}^*\}$, a permuted set of the residuals $\{r_{uk}\}$, permuted separately for each tumor.

(4) Create the permuted marker values $\{x_{uk}^*\}$ using

$$x_{uk}^* = \hat{\theta} + r_{uk}^* \text{ if } u < i_k^* \text{ or } u > j_k^*$$
$$= \hat{\mu} + r_{uk}^* \text{ if } i_k^* \le u \le j_k^*,$$

where

$$\hat{\theta} = \frac{(n - \hat{j}_1 + \hat{i}_1 - 1)\hat{\theta}_1 + (n - \hat{j}_2 + \hat{i}_2 - 1)\hat{\theta}_2}{(n - \hat{j}_1 + \hat{i}_1 - 1) + (n - \hat{j}_2 + \hat{i}_2 - 1)}$$

$$\hat{\mu} = \frac{(\hat{j}_1 - \hat{i}_1 + 1)\hat{\mu}_1 + (\hat{j}_2 - \hat{i}_2 + 1)\hat{\mu}_2}{(\hat{j}_1 - \hat{i}_1 + 1) + (\hat{j}_2 - \hat{i}_2 + 1)}.$$

(5) Use the CBS algorithm on the new dataset for each tumor to obtain the estimated endpoints of the regions of allelic change, denoted $(\hat{i}_1^*, \hat{j}_1^*)$ and $(\hat{i}_2^*, \hat{j}_2^*)$. Include the generated data as admissible only if both allelic changes are statistically significant at the 5% level, and if they are concordant, i.e. if $\delta_1^* = \delta_2^*$ where $\delta_1^* = sign(\hat{\mu}_1^* - \hat{\theta}_1^*)$ and $\delta_2^* = sign(\hat{\mu}_2^* - \hat{\theta}_2^*)$, and where $\hat{\mu}_k^*$ and $\hat{\theta}_k^*$ are the respective means of the markers in the altered and normal subsets of the data, respectively.

(6) Calculate the reference value for the test statistic using

$$t^* = \left|\hat{i}_1^* - \hat{i}_2^*\right| + \left|\hat{j}_1^* - \hat{j}_2^*\right|.$$

(7) Repeat the process a large number of times to obtain the distribution of t^*.

The p-value for the resulting test is obtained by determining the number of admissible permutations N_t such that $t^* \le t$. The p-value is then N_t / N, where N is the total number of admissible permutations.

2.3 Validity and Power of the Test Statistic

We have conducted a series of simulations to examine the properties of the test. The first consideration is validity. Is the size of the test less than or equal to the nominal value under the null hypothesis? Recall from the previous section that our testing strategy was designed to offset three validity challenges. The first is that allelic changes do not occur at random, but instead span sites where changes confer a selective advantage. Our test was constructed under the assumption that there exists a mutational hotspot in the region common to the two allelic changes being compared, and the reference distribution is consequently restricted to data configurations in which the admissible intervals of

allelic change must span the hotspot. This aspect of the test cannot be evaluated confidently from a theoretical basis. It is simply an assumption. Logic suggests that this restriction should make the test more conservative, i.e. make the p-values larger, since we are eliminating many potential permutations in all of which at least one of the mutations does not span the hotspot. These permutations should, on average, have test statistics that are larger (i.e. less "close") than the test statistics of admissible permutations. Later, in Section 2.4, we address this assumption indirectly in an empirical evaluation of the size of the test using data from tumors from different patients.

The remaining two validity challenges, the dependence on the CBS algorithm to identify the allelic changes given its preference for selecting small intervals, and the dependence of the test on an estimated signal to noise ratio, are evaluated in the following manner. Data for each marker were assumed to be normally distributed from distributions with a predefined signal strength. That is we specified the mean value for markers at normal copy number, denoted θ, and the mean in the region of allelic change, denoted μ, with common variance σ^2. These were chosen to specify the signal strength, represented by $|\mu - \theta| / \sigma$, and one of the means was set to 0 and the variance set to 1 without loss of generality. For each simulation we first select a true mutational hotspot at marker h. This was randomly generated from the n markers for each data set. We then generated a data set as follows. First the "true" endpoints of the allelic changes were randomly generated, i_1 and i_2 as U$(1, h)$, and j_1 and j_2 as U(h, n). Observed marker values were generated as normal random variables. That is, x_{uk} was generated as N(θ_k, σ^2) for $u < i_k$ or $u > j_k$ and as N(μ_k, σ^2) for $i_k \le u \le j_k$. The CBS algorithm was used on these data to estimate the endpoints, denoted $\hat{i}_1, \hat{j}_1, \hat{i}_2, \hat{j}_2$, and the test statistic t was calculated using (2.1).

Following the procedure outlined in Section 2.2, the p-value was calculated with 1000 replicates from the reference distribution. The entire process was then repeated 1000 times and the relative frequency with which the test was significant is reported as the test size. Thus the simulation standard error is about ± 0.02 in the estimated test size.

The embedded CBS algorithm uses a default of 10,000 permutations to compute its p-value at its default significance level of 0.01. In our case we are using α=0.05, and so we reduced the number of CBS replicates to 2000 for the original statistic t and to 400 for the permuted statistic t^* to minimize the computational burden. Since the simulation relies on the ability of the CBS algorithm to uncover significant intervals, when signal strength or the number of markers is small it may be difficult to create enough significant, concordant draws from the reference distribution. The procedure is allowed as many attempts as necessary to complete the 1000 replicates required, and likewise it is allowed as many attempts as necessary to generate a significant, concordant data set. The results are presented in Table 1 for signal strengths ranging from 0.5 to 3 standard deviations, and for numbers of markers ranging from 65 to 141. The results show that in general the test is valid.

For the bacterial artificial chromosome (BAC) arrays we have been using the numbers of markers per chromosome arm ranged from 20 to 141. However, in our simulations, for arms with relatively few markers the CBS algorithm rarely detected any significant alterations, and so our test has very low power in this setting. These data are from the Spectral Genomics SpectralChip 2600, an array with 2621 markers. The numbers of markers in the simulations are representative of the counts from the chromosome arms on that array. The SpectralChip is typical in size for a modern BAC array. Oligonucleotide arrays consisting of thousands of markers per arm are also in use (Pinkel and Albertson, 2005).

We evaluated the power of the test to detect clonal allelic changes in a similar manner with the exception that the underlying "true" allelic change was assumed to be identical for both tumor 1 and tumor 2. That is we first randomly generated the hotspot at marker h. Then we randomly generated the endpoints of the allelic change below and above the hotspot, i from U$(1, h)$, and j from U(h, n), and set $i_1 = i_2 = i$ and $j_1 = j_2 = j$. The simulations were then generated as outlined above. We also explored whether the location and the length of the allelic change have an impact on power, by repeatedly generating data for allelic changes of a specific length and position. Thus, for

example, we selected a change that began at marker 5 and ended at marker 10 by generating data from this configuration repeatedly and examining the relative frequency with which the test was significant. The results of these power calculations are in Table 2.

The first row represents simulations in which lengths and locations of the true common signals varied randomly. The other rows represent power to detect a change with a specific length and location. Several conclusions may be drawn from the results. First, the granularity of the array is important, i.e. a chromosome arm with many markers is more likely to detect a clonal change than an arm with relatively few markers (for a given signal strength). Second, not surprisingly, the signal strength is important, with higher signal strength leading to greater statistical power. The results also suggest that short allelic changes may have somewhat less power than larger allelic changes. This can be explained as follows. For a given value of μ the non-centrality parameter for identifying an allelic change by the CBS algorithm is parabolic with respect to the width of the allelic change, with a maximum when the true allelic change is exactly half of the width of the chromosome arm. Consequently smaller true intervals lead to a greater relative frequency of CBS false positives which in turn introduce more noise into the comparison of the observed allelic changes in the two tumors when the true change is clonal. In general, arrays with 100 or more markers appear to have high power to detect a clonal change if the difference in means is 2 standard deviations or greater.

The power we can expect from these experiments clearly depends heavily on the signal strength. We have examined the signal strengths empirically in the series of 38 lymphoma patients. For each chromosomal arm with a detected simple allelic change we calculated the standardized mean difference between the markers in the region of allelic change and the remaining markers, and these estimates are plotted in Figure 3, based on a sample size of 111 observed allelic changes. The results suggest that the signal strength is typically quite strong, averaging about 2 standard deviations.

2.4 Data Analyses

The preceding simulated calculations of size and power are compromised by the fact that they are embedded with some key assumptions, notably the assumption that observed concordant allelic changes necessarily span a mutational hotspot, and the assumption that marker values have normally distributed errors. To examine the validity of the test from an empirical perspective that does not depend on these assumptions we have reanalyzed the dataset of 38 tumors from patients with diffuse large B-cell lymphoma (Chen et al. 2006). For each chromosomal arm (excluding the X chromosome) we compared the profiles for every pair of patients, and performed the test for any observed concordant changes. Since the tumors came from different patients none of the observed concordant changes are clonal, by definition. Thus, the proportion of tests that are significant should approximate the nominal 5% level. In this exercise we identified a total of 745 pairs of changes that were individually detected as significant by the CBS algorithm on the same chromosome arm. These yielded 42 statistically significant comparisons using our algorithm, for an empirical test size of 0.056 (42/745).

We also have access to an additional set of matched ACGH results from two tumors in each of two patients. Both of these examples are of tumors that are presumptively clonal. The tumors from the first patient are both mediastinal metastases (the "mediastinal" patient). The tumors of the other patient are of a primary testis cancer, matched with a lymph node metastasis (the "testis" patient). The complete arrays for the mediastinal patient are displayed in Figure 1. Visual examination of the arrays from the two tumors indicates that there are plausibly clonal concordant changes on arms 2, 5, 6, 7, 24, 34 and 39. The test statistics for these 7 comparisons are 1, 0, 9, 2, 1, 3, and 0, respectively, based on numbers of markers 94, 70, 93, 46, 84, 43 and 35. These correspond to, respectively, p-values of 0.010, 0.010, 0.095, 0.035, 0.005, 0.135, and 0.005. The collective levels of significance of these changes provide strong evidence of clonality in that 5 of the changes are nominally significant, and four are quite highly significant. Data from these arms are displayed with greater magnification in Figure 4, along with reference distributions for the test statistics. The testis patient presents less

visual evidence for clonality. The array for the first tumor detects only two significant allelic changes, on arms 4 and 16. For the second tumor, changes are detected on arms 4, 20, 22, 34 and 38. Thus arm 4 is the only place where a possible clonal event may have occurred. This arm has 141 markers, the test statistic for the comparison is 3 and the corresponding p-value is 0.015.

3. DISCUSSION

The results demonstrate that the proposed test can deliver quite high power for establishing the clonality of observed concordant allelic changes on the two tumors. However, to achieve high power, in excess of 90%, the data from the relevant chromosomal arm must be sufficiently granular (100 markers or greater) and the signal strength must be sufficiently high (around 2 standard deviations or greater). In the absence of these conditions the CBS algorithm is too imprecise at determining the beginning and the end of the regions of allelic gain or loss to establish clonality with sufficient confidence. However, as array technology develops, both the granularity and the signal strength are likely to increase, promising greater resolution accuracy for this method.

An intangible feature of the method in practice will be the manner in which investigators choose the allelic changes that deserve to be tested for clonality. Investigators are only likely to be inclined to perform the test for allelic changes that look close enough to be clonal. Consequently there is an ill-defined multiple testing aspect to the process of evaluating the totality of available data from the arrays. In general, arrays that exhibit lots of allelic changes are more likely to exhibit some changes that "look" clonal, simply by chance, and so the false positive rate must be higher in this case, though it is not obvious how to formulate this issue precisely.

In practice, the ultimate aim is to determine if the tumors are clonal, and this should involve assessing the evidence for and against clonality across the entire genome.

This might involve conducting tests for all of the simple or partial allelic changes that are plausibly clonal upon inspection, but the issue of how to combine this evidence into a unitary decision rule is challenging. For example, suppose that we test three locations where allelic changes look similar on the two tumors. Does the presence of three statistically significant tests with p-values only modestly lower than the nominal 5% level provide more convincing evidence for clonality than, say, a result in which one of the tests is very highly significant but the other two are non-significant? In general, if there are many somatic changes that are clearly independent should this diminish our confidence in the clonal hypothesis in the face of a single, highly significant match? This question rests on the issue of how common it is to find clonal tumors that exhibit lots of bystander (independent) allelic changes. We feel that these questions are difficult to address at present, without access to much more data on the patterns of mutations in clonal and independent pairs of tumors. The proposed test is offered as a building block for interrogating the full genomic evidence available from an individual patient, and as a framework for future research on the most appropriate statistical techniques for addressing this important problem.

Acknowledgements

We are grateful to Jim Korkola and Raju Chaganti for advice on the manuscript, and to Jim Korkola for permission to use his data. The research was supported by the National Cancer Institute, Award CA098438.

References

Begg, C.B., Eng K, Hummer AJ. (2006). Statistical tests for clonality. Biometrics (in press).

Brinkmann, D., Ryan, A., Ayhan, A., McCluggage, W.G., Feakins, R., Santibanez-Korf, M.F. et al. (2004). A molecular genetic and statistical approach for the diagnosis of dual-site cancers. J Natl Cancer Inst 96, 1441-1446.

Chen, W., Houldsworth, J., Olshen, A.B., Nanjangud, G., Chaganti, S., Venkatraman, E.S., Halaas, J., Teruya-Feldstein, J., Zelenetz, A.D., Chaganti, R.S. (2006). Array comparative genomic hybridization reveals genomic copy number changes associated with outcome in diffuse large B-cell lymphomas. Blood 107:2477-2485.

Chunder, N., Roy, A., Roychoudhury, S., and Panda, C. K. (2004). Molecular study of clonality in multifocal and bilateral breast tumors. Pathol Res Pract 200:735-741.

Dacic, S., Ionescu, D. N., Finkelstein, S., and Yousem, S. A. (2005). Patterns of allelic loss of synchronous adenocarcinomas of the lung. Am J Surg Pathol 29:897-902.

Geurts, T. W., Nederlof, P. M., van den Brekel, M. W., van't Veer, L. J., de Jong, D., Hart, A. A., van Zandwijk, N., Klomp, H., Balm, A. J., and van Velthuysen, M. L. (2005). Pulmonary squamous cell carcinoma following head and neck squamous cell carcinoma: metastasis or second primary? Clin Cancer Res 11:6608-6614.

Goldstein, N. S., Vicini, F. A., Hunter, S., Odish, E., Forbes, S., and Kestin, L. L. (2005a). Molecular clonality relationships in initial carcinomas, ipsilateral breast failures, and distant metastases in patients treated with breast-conserving therapy: evidence suggesting that some distant metastases are derived from ipsilateral breast failures and that metastases can metastasize. Am J Clin Pathol 124:49-57.

Goldstein, N. S., Vicini, F. A., Hunter, S., Odish, E., Forbes, S., Kraus, D., and Kestin, L. L. (2005b). Molecular clonality determination of ipsilateral recurrence of invasive breast carcinomas after breast-conserving therapy: comparison with clinical and biologic factors. Am J Clin Pathol 123:679-689.

Hiroshima, K., Toyozaki, T., Kohno, H., Ohwada, H., and Fujisawa, T. (1998). Synchronous and metachronous lung carcinomas: molecular evidence for multicentricity. Pathol Int, 48: 869-876.

Holst, V. A., Finkelstein, S., and Yousem, S. A. (1998). Bronchioloalveolar adenocarcinoma of lung: monoclonal origin for multifocal disease. Am J Surg Pathol 22:1343-1350.

Huang, J., Behrens, C., Wistuba, I., Gazdar, A. F., and Jagirdar, J. (2001). Molecular analysis of synchronous and metachronous tumors of the lung: impact on management and prognosis. Ann Diagn Pathol 5:321-329.

Imyanitov, E. N., Suspitsin, E. N., Grigoriev, M. Y., Togo, A. V., Kuligina, E., Belogubova, E. V., Pozharisski, K. M., Turkevich, E. A., Rodriquez, C., Cornelisse, C. J., Hanson, K. P., and Theillet, C. (2002). Concordance of allelic imbalance profiles in synchronous and metachronous bilateral breast carcinomas. Int J Cancer 100:557-564.

Janschek, E., Kandioler-Eckersberger, D., Ludwig, C., Kappel, S., Wolf, B., Taucher, S., Rudas, M., Gnant, M., and Jakesz, R. (2001). Contralateral breast cancer: molecular differentiation between metastasis and second primary cancer. Breast Cancer Res Treat 67:1-8.

Jiang, J.K., Chen, Y.J., Lin, C.H., Yu, I.T. and Lin, J.K. (2005). Genetic changes and clonality relationship between primary colorectal cancers and their pulmonary metastases--an analysis by comparative genomic hybridization. Genes Chromosomes Cancer 43:25-36.

Kollias, J., Man, S., Marafie, M., Carpenter, K., Pinder, S., Ellis, I. O., Blamey, R. W., Cross, G., and Brook, J. D. (2000). Loss of heterozygosity in bilateral breast cancer. Breast Cancer Res Treat 64:241-251.

Kuukasjarvi, T., Karhu, R., Tanner, M., Kahkonen, M., Schaffer, A., Nupponen, N., et al. (1997). Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. Cancer Res 57, 1597-1604.

Lai, W.R., Johnson, M.D., Kucherlapati, R. and Park, P.J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. Bioinformatics 21, 3763-70.

Lau, D. H., Yang, B., Hu, R., and Benfield, J. R. (1997). Clonal origin of multiple lung cancers: K-ras and p53 mutations determined by nonradioisotopic single-strand conformation polymorphism analysis. Diagn Mol Pathol 6:179-184.

Leong, P. P., Rezai, B., Koch, W. M., Reed, A., Eisele, D., Lee, D. J., Sidransky, D., Jen, J., and Westra, W. H. (1998). Distinguishing second primary tumors from lung metastases in patients with head and neck squamous cell carcinoma. J Natl Cancer Inst 90:972-977.

Matsuzoe, D., Hideshima, T., Ohshima, K., Kawahara, K., Shirakusa, T., and Kimura, A. (1999). Discrimination of double primary lung cancer from intrapulmonary metastasis by p53 gene mutation. Br J Cancer 79:1549-1552.

Murase, T., Takino, H., Shimizu, S., Inagaki, H., Tateyama, H., Takahashi, E., Matsuda, H., and Eimoto, T. (2003). Clonality analysis of different histological components in combined small cell and non-small cell carcinoma of the lung. Hum Pathol 34:1178-1184.

Olshen, A. B., Venkatraman, E. S., Lucito, R., and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics 5:557-572.

Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.,L., Chen, C., Zhai, Y., Dairkee, S. H., Ljung, B. M., Gray, J. W., and Albertson, D.G. (1998). High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet 20:207-211.

Pinkel, D. and Albertson, D. G. (2005). Comparative genomic hybridization. Ann Rev Genomics Hum Genet 6:331-354.

Regitnig, P., Ploner, F., Maderbacher, M., and Lax, S. F. (2004). Bilateral carcinomas of the breast with local recurrence: analysis of genetic relationship of the tumors. Mod Pathol 17:597-602.

Schlechter, B.L, Yang, Q., Larson, P.S., Golubeva, A., Blanchard, R.A, de las Morenas, A. and Rosenberg, C. L. (2004). Quantitative DNA fingerprinting may distinguish new primary breast cancer from disease recurrence. J Clin Onc 22, 1830-1838.

Shimizu, S., Yatabe, Y., Koshikawa, T., Haruki, N., Hatooka, S., Shinoda, M., Suyama, M., Ogawa, M., Hamajima, N., Ueda, R., Takahashi, T., and Mitsudomi, T. (2000). High frequency of clonally related tumors in cases of multiple synchronous lung cancers as revealed by molecular diagnosis. Clin Cancer Res 6:3994-3999.

Shin, S. W., Breathnach, O. S., Linnoila, R. I., Williams, J., Gillespie, J. W., Kelley, M. J., and Johnson, B. E. (2001). Genetic changes in contralateral bronchioloalveolar carcinomas of the lung. Oncology 60:81-87.

Sieben, N. L.G. , Kolkman-Uljee, S.M., Flanagan, A.M., le Cessie, S., Cleton-Jansen, A.M., Cornelisse, C.J., and Fleuren, G.J. (2003). Molecular genetic evidence for monoclonal origin of bilateral ovarian serous borderline tumors. Am J Pathol 162:1095-1101.

Sozzi, G., Miozzo, M., Pastorino, U., Pilotti, S., Donghi, R., Giarola, M., De Gregorio, L., Manenti, G., Radice, P., Minoletti, F. (1995). Genetic evidence for an independent origin of multiple preneoplastic and neoplastic lung lesions. Cancer Res 55:135-140.

Stenmark-Askmalm, M., Gentile, M., Wingren, S., and Stahl, O. (2001). Protein accumulation and gene mutation of p53 in bilateral breast cancer. South-East Sweden Breast Cancer Group. Acta Oncol 40:56-62.

Teixeira, M.R., Ribeiro, F.R., Torres, L., Pandis, N., Anderson, J.A., Lothe, R.A. and Heim, S. (2004). Assessment of clonal relationships in ipsilateral and bilateral multiple breast carcinomas by comparative genomic hybridization and hierarchical clustering analysis. Br J Cancer 91, 775-782.

Tse, G. M., Kung, F. Y., Chan, A. B., Law, B. K., Chang, A. R., and Lo, K. W. (2003). Clonal analysis of bilateral mammary carcinomas by clinical evaluation and partial allelotyping. Am J Clin Pathol 120:168-174.

van Rens, M. T., Eijken, E. J., Elbers, J. R., Lammers, J. W., Tilanus, M. G., and Slootweg, P. J. (2002). p53 mutation analysis for definite diagnosis of multiple primary lung carcinoma. Cancer 94:188-196.

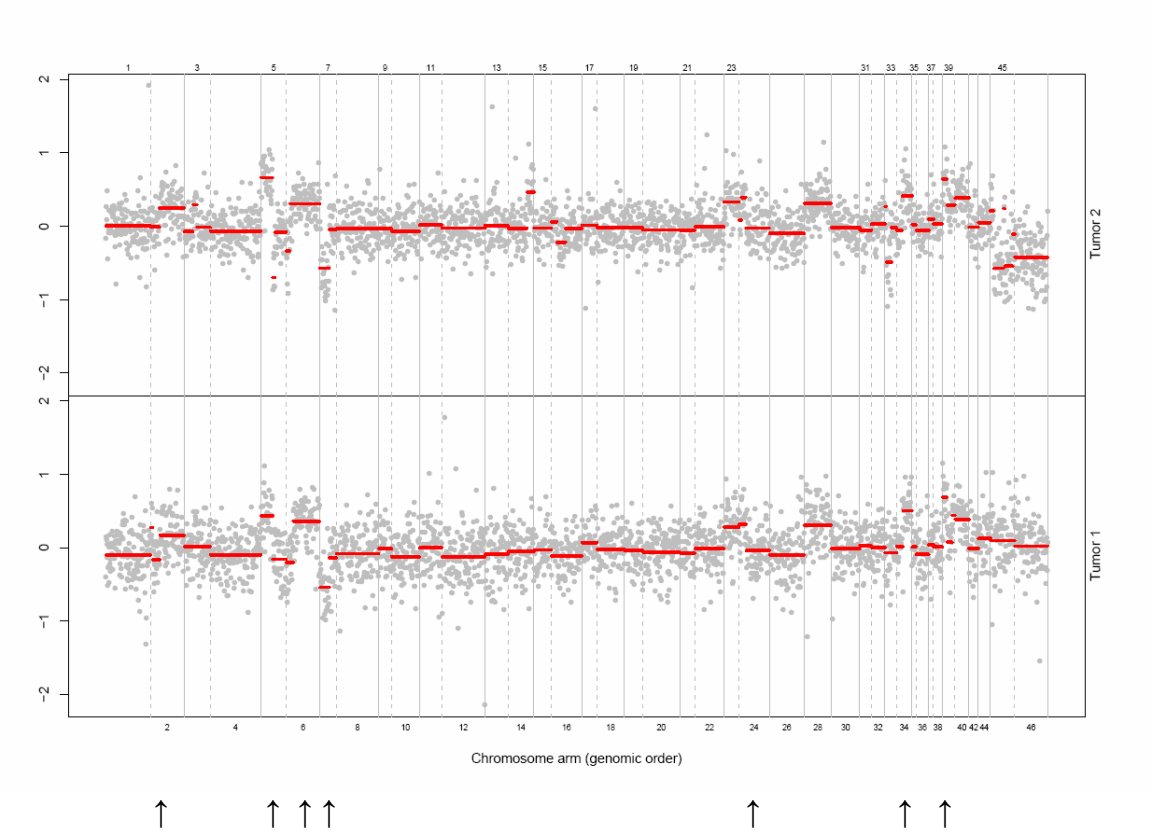


Figure 1. Log ratio measurements from ACGH of two different metastases from the "mediastinal patient." The horizontal lines show CBS estimated segments, and the vertical lines separate chromosomal arms (1 to 46 representing chromosomes 1-22 and chromosome X). The arrows indicate chromosome arms where plausible clonal changes may have occurred.

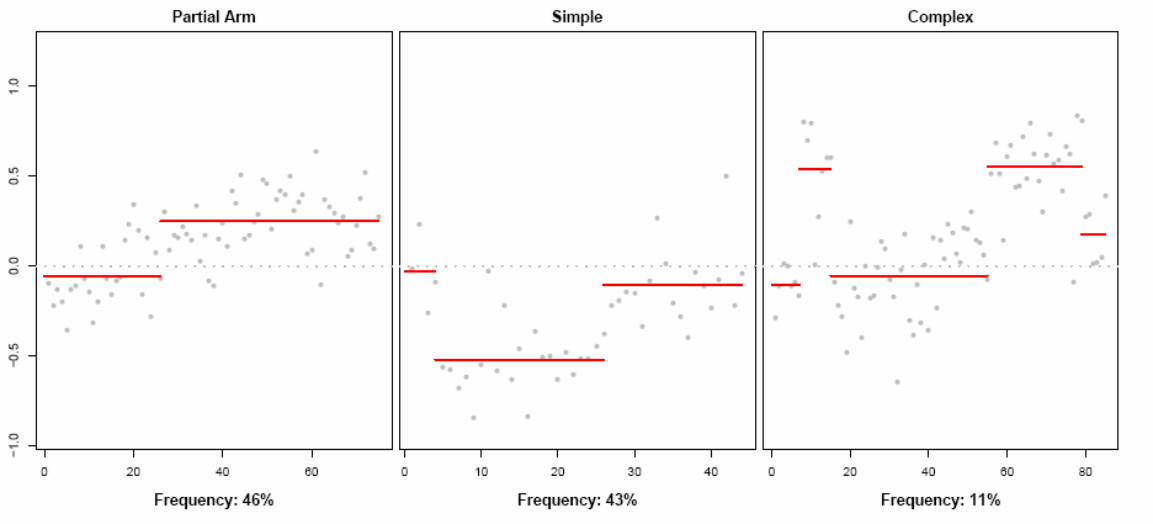


Figure 2. Data examples of alteration patterns. The dotted line indicates normal copy number. The frequencies of these patterns were determined empirically from a sample of 38 independent patients with cancer.

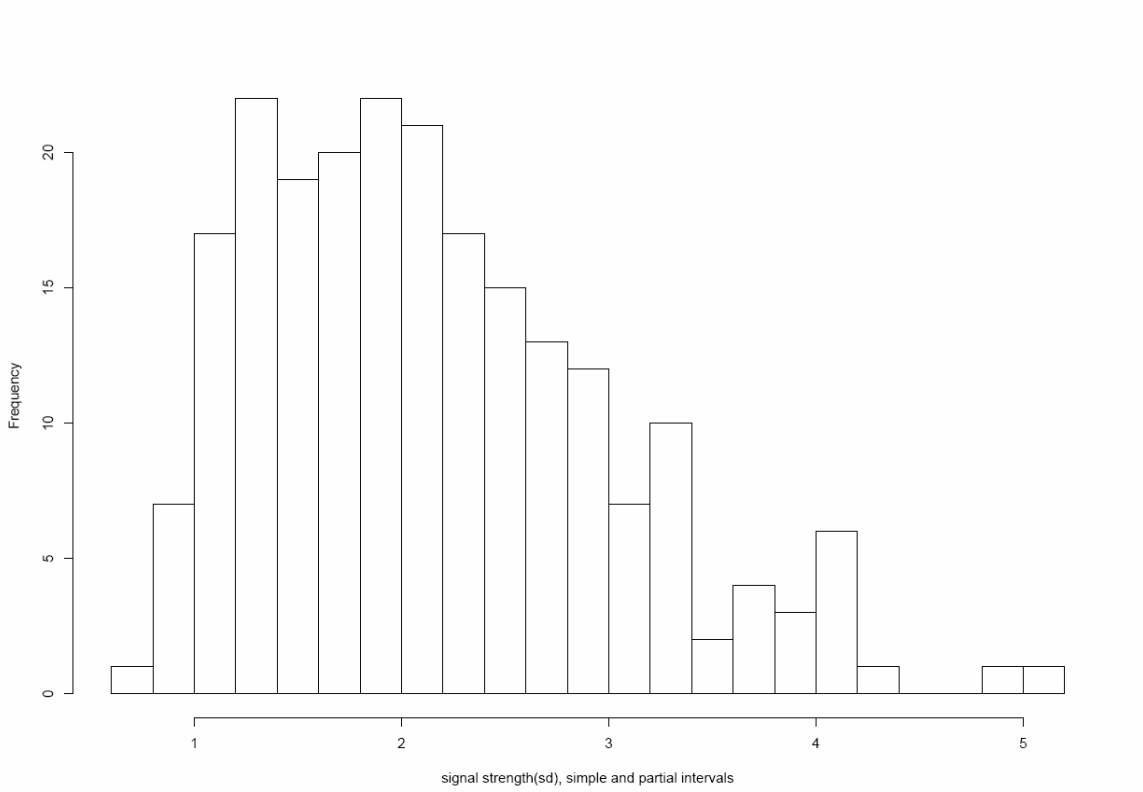
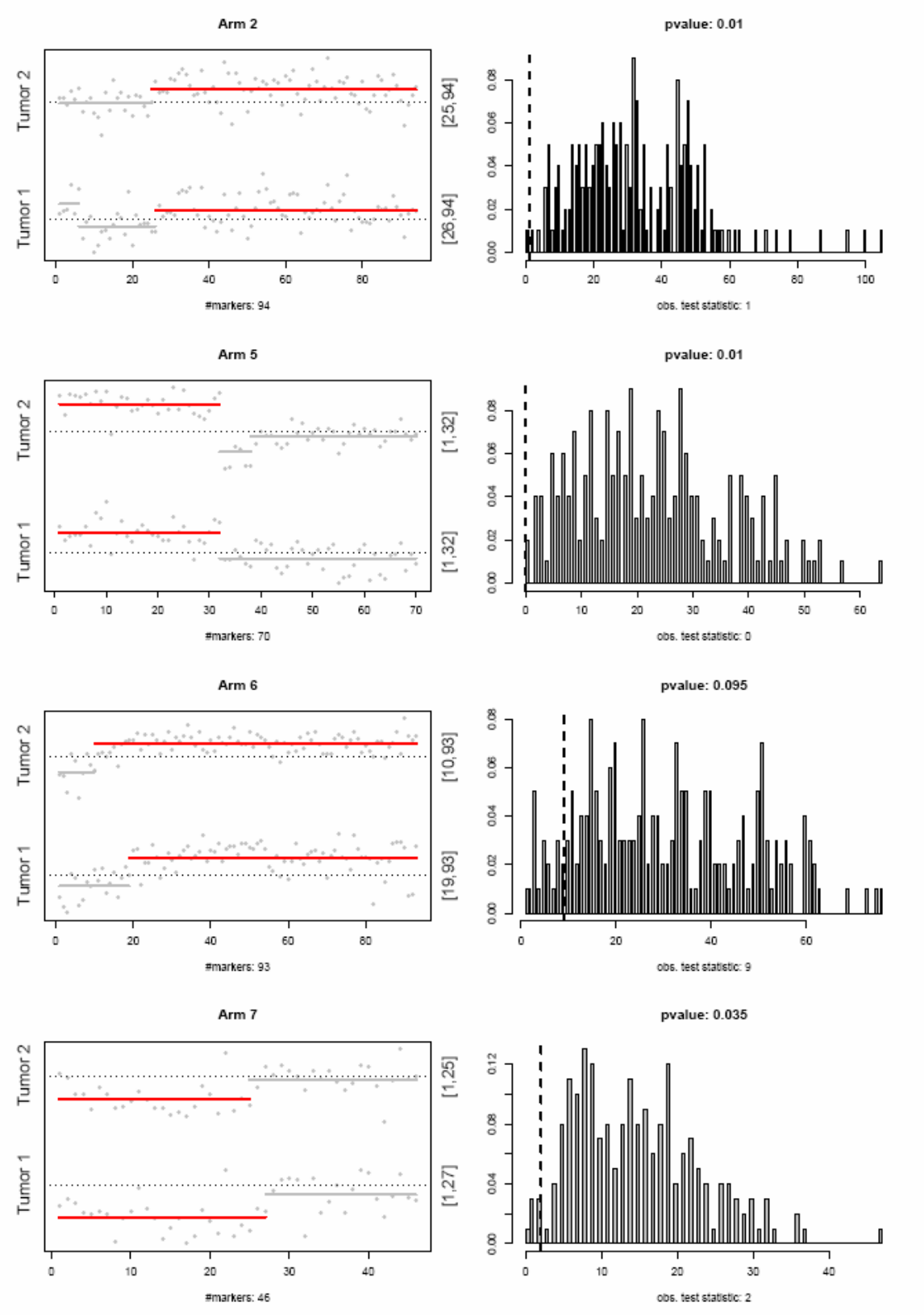


Figure 3. Histogram of estimated signal strengths derived empirically from 38 patients with lymphoma.



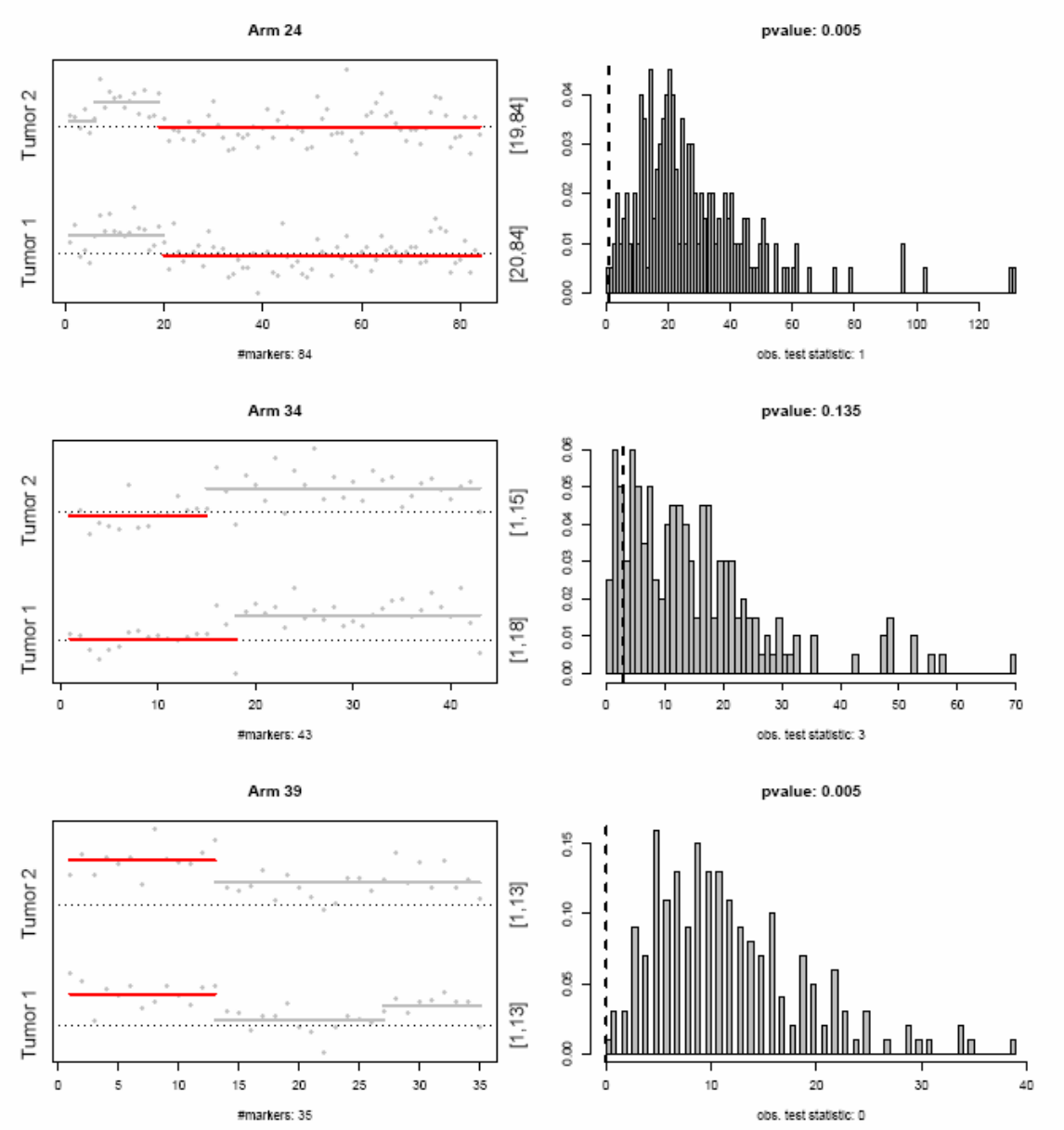


Figure 4. Details of chromosomal arms and corresponding clonality test permutation reference distributions for the concordant changes in the "mediastinal" patient. On the histogram, the vertical dotted line denotes the observed test statistic, and the proportion equal to and to the left of the line determines the p-value.

Table 1. *Simulated size of the clonality test measured as the proportion of significant tests (p <=0.05) in 1000 pairs using randomly selected concordant intervals spanning a common hotspot.*

# Markers	Signal Strength (S.D.)			
	0.5	**1.0**	**2.0**	**3.0**
65	0.02	0.05	0.05	0.06
100	0.04	0.04	0.05	0.06
141	0.05	0.05	0.06	0.05

Table 2. *Simulated power of the clonality test for various signal strengths and alternatives.*

# Markers	65		100		141	
Signal Strength	**1**	**2**	**1**	**2**	**1**	**2**
Random	0.29	0.85	0.49	0.94	0.60	0.98
[1,5]	0.31	0.78	0.40	0.87	0.38	0.90
[5,10]	0.27	0.72	0.37	0.80	0.39	0.88
[1,30]	0.38	0.87	0.50	0.95	0.62	0.98
[5,35]	0.23	0.86	0.38	0.93	0.52	0.98
[30,60]	0.30	0.83	0.47	0.97	0.59	0.98