Testing Clonal Relatedness of Tumors Using Array Comparative Genomic Hybridization: A Statistical Challenge

Irina Ostrovnaya and Colin B. Begg

Abstract

In recent years many investigators have sought to use array technologies that characterize somatic alterations in tumors, such as array comparative genomic hybridization (ACGH), to classify pairs of tumors from the same patients as either independent primary cancers or metastases. A wide variety of strategies have been proposed. Several groups have endeavored to use hierarchical clustering for this purpose. This technique was popularized in genomics as a means of finding clusters of patients with similar gene expression patterns with a view to finding subcategories of tumors with distinct clinical characteristics. Unfortunately, this method is not well suited to the problem of classifying individual pairs of tumors as either clonal or independent. In this article we show why hierarchical clustering is unsuitable for this purpose, and why this method has the paradoxical property of producing a declining probability that clonal tumor pairs will be correctly identified as more information is accrued (i.e., more patients). We discuss alternative strategies that have been proposed, which are based on more conventional conceptual formulations for statistical testing and diagnosis, and point to the remaining challenges in constructing valid and robust techniques for this problem.

The advent of powerful new technologies in molecular genetics has allowed for the opportunity to characterize tumors in tremendous detail at the molecular level. An important potential clinical ramification of these developments is the opportunity for more accurate diagnosis of new tumors. In particular, when a new tumor develops in survivors of an existing cancer one must determine whether the new tumor represents an independent primary or a metastasis from the initial tumor. If it is a metastasis, then the new tumor is, by definition, clonally related to the initial primary, and as such should display somatic changes that characterize the initial primary.

In recent years many investigators have sought to examine and compare the somatic molecular fingerprints of tumors with a view to using the results to assist in and possibly alter the pathologic diagnosis of the second tumor. Much of this work has focused on profiles of mutations at candidate genetic loci, such as the presence of loss of heterozygosity or individual point mutations that occur commonly in the tumor sites under investigation (1–25). More recently, some investigative groups have sought to tackle this problem by using a comprehensive, genewide strategy (26–44). In this approach, array comparative genomic hybridization (ACGH) or related techniques are used to try to identify the locations of all somatic copy number gains and losses. One then compares the patterns of these gains and losses in the two tumors to see if they are sufficiently similar that we can conclude that the tumors share a clonal origin, i.e., if the cells in the tumors originated from the original clonal cell. Because ACGH arrays consist of so much data, in some cases hundreds of thousands of marker values covering all of the chromosomes, it is a big challenge to organize the data in a way that effectively synthesizes the information relevant to determining whether the tumors are clonal. In fact, the literature on this topic is extraordinarily fragmented, with different investigative groups essentially inventing their own strategies for data summarization and analysis. The purpose of this article is to show that hierarchical clustering, a statistical framework that has been used by many investigators in this context, is an inappropriate technique for this purpose. We also explain how to construct the problem in a more appropriate, hypothesis testing framework, and provide comparative analyses of data from patients with lung cancer to illustrate the concepts.

Data Configurations

Array copy number data consist of a set of continuous marker values arranged consecutively across the entire genome. The data are preprocessed in such a way that the mean value of the markers (usually set to zero) represents the normal copy number (two alleles). In normal tissues
Testing for Clonal Relatedness of Tumors

Translational Relevance

The pathological diagnosis of a new tumor in a patient with an existing or previous cancer can be challenging for pathologists in certain clinical settings. Modern techniques for molecular analysis of the two tumor specimens to detect common somatic changes offer the promise of more accurate classification of the second tumor as either a metastasis or a second independent primary. Analyses of the tumors for genome-wide copy number changes using array CGH technology has been used for this purpose in a number of research settings. However, the statistical analysis is conceptually challenging. This article helps to clarify the conceptual issues, and lays the groundwork for the development of valid analytic techniques for improved diagnoses in future clinical applications.

the marker values should thus represent random deviations around the mean value representing two copies, except for occasional very short regions that represent germ-line variations in copy number. By contrast, tumors consist typically of many regions that are characterized by gains or losses of alleles, with these regions of gains and losses varying in length. In Fig. 1 we see an array from a tumor of a patient with lung cancer. The red lines above the zero line represent regions where the data indicate gains, and the blue lines below zero represent losses. Some of these gains and losses represent the initial, clonal mutational events that provided the cells with the growth advantage that led to the development of the cancer, whereas others may be subsequent events that occurred in the milieu of genetic instability created by the tumor. These gains and losses characterize the tumor's somatic genetic fingerprint, and the patterns should share common features in tumors that originate from the same clonal cell. Notice that the segmentation in Fig. 1 is relatively clear cut. That is, it is obvious visually where the gains and losses are located. Frequently arrays contain more statistical noise, and it is harder to identify regions of gain and loss with precision.

The statistical challenge is to compare arrays from each of two tumors from the same patient to see whether the patterns of allelic changes seem sufficiently similar to suggest that the tumors share a clonal origin. A data set will typically consist of tumor pairs from a number (N) of patients from the clinical scenario under investigation. This could be women with bilateral breast cancer, or pairs of lung tumors that may or may not be independent primaries, or lung tumors paired with previously diagnosed head and neck primaries, etc. Consequently the data set will consist of 2N tumors, with the tumors being paired by patient, and with a large number of markers characterizing each tumor. The goal of the analysis is to determine, for each tumor pair, whether the tumors are clonal or independent. More generally, we wish to determine an algorithm that will allow diagnosis of tumor pairs from future patients that are not part of the initial data set.

Hierarchical Clustering

Hierarchical clustering is a statistical method to discover and visualize structure in a multivariate data set. It builds a tree-like model in which different data objects (e.g., tumors in our case) or clusters of objects are connected to each other in order of their degree of similarity. The similarity between two tumors is often measured by the correlation coefficient of the marker values in the pair of tumors or alternatively by the Euclidean distance between them. The similarity between two clusters of tumors is determined by the linkage function. This could be the distance between the two closest (single linkage) or farthest (complete linkage) members of the clusters, or the average distance across all pairs of members of the two clusters (average linkage). In the clonality literature, two tumors from a patient are considered to be of clonal origin if they cluster together in a terminal branch of the tree.

There have been many recent articles published that exploit hierarchical clustering in this setting. This includes work by Waldman and colleagues (36), Texiera and colleagues (26), Wong and colleagues (41), Torres and colleagues (39), Ghazani and colleagues (37), Agelopoulos and colleagues (38), Brommesson and colleagues (42), Liu and colleagues (43), and Yang and colleagues (44). Although many of these authors also used other analytical and biological methods to support their conclusions, hierarchical clustering was one of the principal tools. Several of the preceding articles explicitly state or assume that tumors that cluster together in a terminal branch are closer to each other than to any other tumor with respect to the measure of similarity used in the analysis. However, this is not true unless single linkage is used. Instead, to our knowledge, studies of clonality have used complete, average, or Ward linkage (45), and none of them used single linkage. Thus, the diagnosis for any two tumors from a particular patient is influenced by both their relationship to other tumors in the cohort and the clustering patterns of other tumors among themselves. Yet the diagnosis of clonality in an individual patient should be informed primarily by the similarity of the tumors from the patient under investigation, rather than by their relationships with tumors from other patients.

In our view, hierarchical clustering is an inappropriate analytic tool for assessing tumor clonality. Hierarchical clustering is an unsupervised classification technique that assumes the number of clusters and their members are unknown, and its aim is to uncover possible general structure in the data. Conversely, in clonality testing the hypothetical clusters (tumors within patients) are known, and the task at hand is to test whether pairs of tumors within a hypothetical cluster are independent or clonal. Omitting this known structure is not an efficient use of information,
and it is naturally more difficult to identify the structure when all tumors are mixed up and pooled together.

On a related note, even if there is absolutely no dependence structure in the data, by design hierarchical clustering still identifies clusters, and by chance there are likely to be tumors from the same patient that end up in a “clonal” pairing. Thus, using hierarchical clustering without any significance assessment may be misleading. There are statistical methods that allow one to test whether clusters are reproducible by assessing similarity between clusters of perturbed data (46). We note that some of the referenced studies, but not all, assessed whether the number of pairs of tumors determined to be clonal is higher than what would be expected by chance. To accomplish this the patient labels are randomly permuted and assigned to the final branches of the dendrogram, and the number of within-patient pairings is counted. Permutations are then repeated many times. If the original number of clonal patients is in the upper fifth percentile of the counts from these permutations then there is statistically significant evidence of genuine clustering in the data. Some other similar techniques have also been used. Although such an approach can test for false positive clustering overall in the data, it cannot assess the confidence of the diagnosis for any particular patient.

Finally, in using hierarchical clustering the more tumors there are in the data set, the less likely it will be that any two particular clonal tumors will be sufficiently close to each other to cluster together in a terminal branch and be classified as clonal. This result occurs simply because, as the data set increases, the clonal pair has to beat more and more competitors. Thus, inevitably, the sensitivity of hierarchical clustering to detect specific clonal pairs must decrease steadily as the sample size increases. These concepts are illustrated later in the section entitled “Simulation.”

Conventional Paradigms of Statistical Diagnosis and Testing

The data structure we face in clonality testing fits much more naturally into a familiar statistical hypothesis or diagnostic testing framework. In this setting, as in hierarchical clustering, one still needs to select a measure that characterizes the similarity of any pair of tumors. In the next section we describe some strategies for choosing a good similarity measure. However, the conceptual structure for addressing the diagnosis of tumor pairs differs markedly from hierarchical clustering once the measure is chosen. In hypothesis or diagnostic testing we need to distinguish between two hypotheses: the hypothesis that the tumors arose independently (the “null” or “independence” hypothesis), and the hypothesis that the two tumors are clonally related (the “alternative” or “clonal” hypothesis). These statistical hypotheses correspond exactly to the two clinical diagnoses that the clinician or pathologist wants to distinguish: a new independent primary versus a metastasis. Moreover, these are not global hypotheses that encompass the entire data set. Rather, they represent individual diagnoses that we need to make for each patient. That is we need to classify whether each one of the patient-specific tumor pairs is independent or clonal.

In traditional statistical significance testing we approach this classification by focusing primarily on one of the two hypotheses, usually termed the null hypothesis. In our context this is most conveniently the hypothesis that the tumors are independent. We determine if the similarity measure for the tumors from the patient under consideration is representative of the distribution of similarity measures from independent tumors. This is a convenient strategy in our context because it is easy to characterize the reference distribution for the similarity measure among...
independent tumor pairs. We can accomplish this by studying pairings of tumors from different patients, because these cannot possibly be clonally related. Thus in a data set of $N$ patients, each with two tumors, we have $N(N-1)/2$ independent pairs of patients that can be used to create this reference distribution, and four tumor pairings per patient pair, leading to $2N(N-1)$ independent tumor pairs. The similarity measure obtained by comparing the two tumors from the same patient can then be placed in this reference distribution, and if the measure is statistically significant (e.g., in the top, say, fifth percentile) the patient’s tumors are deemed to be clonal, and if the measure is not statistically significant, the patient’s tumors are deemed to be independent. Unlike hierarchical clustering, this method does not lose power as the number of patients increases, and in fact the reference distribution becomes known with greater precision. In addition, the type I error can be specified at the discretion of the investigator. This approach provides a level of significance for each individual patient, a goal that is not possible to achieve via hierarchical clustering.

**Simulation**

Earlier, we made the argument as a statement of principle that hierarchical clustering will have poor statistical properties in the context of clonality testing. In particular we stated that the probability of detecting truly clonal tumor pairs will necessarily decline as the sample size increases. We also conjectured that hierarchical clustering must be inefficient because it is not constructed in recognition of the inherent paired structure of the data. In this section we present the results of a small but illustrative simulation study to show these principles in action.

For simplicity of exposition, the study is a very simple cartoon of the real setting, but it contains the essential conceptual features. We use arrays with 10 markers (usually these contain several thousand), and clonality is represented simply by the presence of correlation between marker values in the pair of tumors under investigation; independence is represented by a correlation of 0. For independent tumor pairs the marker values are generated as normal random variables with mean 0 and variance 1. For clonal tumor pairs, the pairs of corresponding marker values in the two tumors are generated as independent bivariate normal variables each with mean 0, variance 1, and correlation 0.8. Patients are independent of each other. We generate a data set of varying numbers of patients, and varying proportions of clonal and independent tumors. The Pearson correlation coefficient serves as the similarity measure. For each generated data set we analyzed the data using hierarchical clustering with average linkage, because this is the technique commonly used, and counted how many clonal and independent tumors clustered together and were declared clonal. For the hypothesis testing approach, we created a reference distribution by calculating correlation coefficients of all possible pairings of tumors from different patients. For each original patient the $P$ value was determined with respect to this reference distribution. We generated 10,000 data sets in this manner and averaged the results, which are presented in Table 1.

The columns headed “Hierarchical Clustering” show clearly that both the false positive and true positive rates for detecting correlated (clonal) pairs decline steadily as

<table>
<thead>
<tr>
<th>No. clonal pairs</th>
<th>No. independent pairs</th>
<th>Hierarchical clustering</th>
<th>Hypothesis testing</th>
<th>Calibrated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>True positive rate</td>
<td>False positive rate</td>
<td>True positive rate</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>0.833</td>
<td>0.052</td>
<td>0.929</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>0.779</td>
<td>0.030</td>
<td>0.929</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>0.680</td>
<td>0.014</td>
<td>0.929</td>
</tr>
<tr>
<td>5</td>
<td>50</td>
<td>0.598</td>
<td>0.007</td>
<td>0.929</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>0.800</td>
<td>0.038</td>
<td>0.929</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.751</td>
<td>0.025</td>
<td>0.927</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>0.669</td>
<td>0.012</td>
<td>0.927</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>0.582</td>
<td>0.007</td>
<td>0.928</td>
</tr>
<tr>
<td>25</td>
<td>5</td>
<td>0.723</td>
<td>0.021</td>
<td>0.928</td>
</tr>
<tr>
<td>25</td>
<td>10</td>
<td>0.692</td>
<td>0.016</td>
<td>0.928</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>0.628</td>
<td>0.009</td>
<td>0.928</td>
</tr>
<tr>
<td>25</td>
<td>50</td>
<td>0.557</td>
<td>0.006</td>
<td>0.928</td>
</tr>
<tr>
<td>50</td>
<td>5</td>
<td>0.640</td>
<td>0.012</td>
<td>0.928</td>
</tr>
<tr>
<td>50</td>
<td>10</td>
<td>0.622</td>
<td>0.009</td>
<td>0.928</td>
</tr>
<tr>
<td>50</td>
<td>25</td>
<td>0.578</td>
<td>0.007</td>
<td>0.929</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>0.523</td>
<td>0.005</td>
<td>0.928</td>
</tr>
</tbody>
</table>
the sample size increases. This result is really not surprising because it must become harder for a correlated pair to succeed in connecting in a terminal branch of the tree as it competes with more and more alternative pairings. In contrast, the true positive rates for the “Hypothesis Testing” strategy are largely unaffected by either the sample size or the proportion of “clonal” pairs in the sample. The overall comparative diagnostic performance of these two approaches is not immediately apparent from these numbers because a low true positive rate tends to be associated with a low false positive rate. To calibrate the comparison, in the last two columns we have adjusted (approximately) the criterion for detecting clonal pairs using hypothesis testing by resetting it to ensure that the false positive rate is similar to the corresponding false positive rate for hierarchical clustering. Comparing the resulting true positive rates (column 3 versus column 7), we see that the hypothesis testing approach consistently outperforms hierarchical clustering by several percentage points, suggesting that this approach is more efficient. We conjecture that this modest efficiency deficit is due to the fact that hierarchical clustering does not recognize the paired nature of the design, and that a similar efficiency loss would occur regardless of the similarity measure used. However, because clonality was represented by simple correlation in both the data generation and the analysis, the results are not informative about how to construct the best measure of similarity in practice, in which the phenomena examined are segmented sections of the genome.

Selection of Diagnostic Measure of Similarity

Regardless of whether we use hierarchical clustering or hypothesis-diagnostic testing as a conceptual framework, the choice of the measure used to characterize similarity between pairs of tumors is of crucial importance. In the following we discuss some of the issues involved in selecting a similarity measure in this context, and express our opinions on its ideal construction, although we acknowledge at the outset that there is no conclusive research available yet that provides a preferred strategy for selecting this measure. Recall that the data consist of pairs of marker values arranged consecutively along the genome. Sometimes the marker values are combined within small regions, chromosome bands or arms, and summarized as +1/0/−1 according to gain-normal-loss status. A basic strategy is to simply characterize the correlation in these pairs of marker values either by using a straightforward measure of correlation, such as Pearson’s correlation coefficient, or by measuring the Euclidian distance between the two arrays of marker values. Indeed some of the articles employing hierarchical clustering have adopted these familiar measures. However, these strategies ignore the underlying biological structure that gives rise to clonal relatedness.

Biological structure is embedded in the patterns of allelic gains and losses in the two tumors, and so the appropriate first step in any analysis involves identification of the locations of these allelic gains and losses on the two tumors, such as those marked by the red and blue lines (respectively) on Fig. 1. This is known as a segmentation analysis, and there are many options available for this purpose. In this article we have used the circular binary segmentation analysis for this purpose, one that is known to have good statistical properties (47).

In the largely accepted model for carcinogenesis, an initial clonal cell gradually accumulates mutational changes that affect the functioning of individual genes in a way that provides a growth advantage to the cell. Many of these mutations are in fact losses or gains of portions of a chromosome arm in which the important genes reside. If the two tumors are clonal, then the second tumor will have
been initiated by the deposit of a metastatic cell from the initial tumor. This cell, being a daughter cell of the initiating clonal cell, must therefore contain the crucial initiating gains and losses that gave rise to the first tumor. A good similarity measure should be constructed to identify these events, which will be represented by closely matching segments on the two tumors. However, the two tumors, even if they share a clonal origin, may, over time, accumulate additional allelic gains and losses that either lead to further growth advantages or are merely bystander mutations. Thus two clonal tumors may have many nonmatching gains and losses, but they must possess at least some gains or losses that are identical.

We believe that the measure of similarity between the two tumors should be based on segmented data, and should reflect the evidence supporting the presence of gains and losses that derive from exact matches. It should also reflect the degree of similarity of changes across the entire genome, i.e., the more frequently closely matching changes are observed, the greater the chance that the tumors are clonal. For example, Waldman and colleagues (36) have used a similarity score that characterizes the broad correlation of chromosomal gains and losses. In this method each chromosome arm is determined to be an allelic gain, a loss, or no change on the basis of the average of all of the markers in the chromosome arm. Individual chromosome arms contribute positively to this score if they exhibit concordant changes, with the contributions from chromosome arms for which allelic changes are infrequent being weighted correspondingly higher. Bollet and colleagues (40) used a partial identity score that is a weighted count of breakpoints that exhibit an exact match on the two tumors, including those at the ends of the chromosome arms, with weights determined by the observed relative frequencies of the breakpoints in the data set. In our own work we have created a measure that takes into account both the correlation of the broad patterns of gains and losses across chromosome arms and the similarity of specific within-arm allelic changes that are plausibly clonal. We denote this as the likelihood ratio (LR) method (48). This method differs importantly from Bollet’s in that the estimated endpoints of these plausibly clonal changes need not match exactly, in recognition of the fact that array segmentation is subject to error. Thus, closeness of the endpoints is judged probabilistically with respect to the chances that these represent true exact matches.

Illustrative Example

We illustrate the concepts using data from a recent study of lung cancer by our group (49). In this study 24 pairs of tumors with available fresh-frozen tissue were identified from the historical tumor bank at Memorial Sloan-Kettering Cancer Center. In all cases the patient had experienced two (in two cases three) lung cancer primaries that had been classified initially as independent tumors using conventional clinico-pathological criteria. In our original analysis we applied our chosen similarity measure (LR), and this led to the diagnosis of 8 of the 24 tumor pairs as either definitely clonal (4 cases with similarity profiles well outside the range of the histogram of all independent pairings of tumors from different patients), or probably clonal (4 cases in the top 95th percentile of this distribution, but not larger than the largest observed independent pairing). The LR values for the 24 tumor pairs comprise the red histogram in Fig. 2, alongside the blue reference histogram consisting of all independent pairings of tumors from distinct patients. In Fig. 3 the segmented arrays for one of the definitely clonal tumor pairs are displayed. In this figure the arrays from the two tumors are aligned vertically. Notice that the arrays are more noisy than the array in Fig. 1, and as a result a number of small allelic changes are detected that do not seem fully convincing, primarily in the lower array, even though they represent
significant changes. Nonetheless there are some very close similarities evident in the segmentation patterns (on chromosomes 1p, 3p, 6q, and 22q), and there is strong evidence that all four of these changes are clonal (due to the closeness of the matches). Analogous or greater degrees of similarity can be observed for the other three definitely clonal cases. Thus, although we have no gold standard diagnosis for any of these patients, the evidence from the four definitely clonal pairings looks visually definitive. Further evidence was available from mutational testing of relatively rare point mutations in genes that are frequently mutated in lung cancer. Eight of the tumor pairs had matching point mutations in either BRAF or KRAS, and these corresponded exactly to the eight pairings classified as either definitely or probably clonal (49). However, we cannot be sure that the remaining 16 pairings are independent simply because no matching mutations in candidate genes were found.

As a comparison to this approach we have analyzed these data using both hierarchical clustering and also the significance testing approach using pairings of tumors from different patients to construct reference distributions. We applied both methods using each of four different similarity measures: correlation of the raw marker values in the arrays; correlation of segmented marker values in the arrays; a similarity measure proposed by Bollet and colleagues (40) termed the partial identity score that consists of a weighted count of allelic changes that occur at exactly the same marker value; and our own preferred similarity measure (LR). The first two measures were selected because they are often used in hierarchical clustering. Note that for all methods we first reduced the dimension (244 K) of the markers by local averaging in groups of 15 markers in order to reduce excessive noise that obscures the segmentation structure. In general, use of full resolution arrays produced inferior results (data are not shown).

The hierarchical clustering dendrogram for the LR similarity measure is displayed in Fig. 4. This can be contrasted with the histograms in Fig. 2 because they are based on the same similarity measure. Notice that six tumor pairings are clustered together in terminal branches. These include the four definitely clonal pairings (numbers 7, 9, 6, 13), and one of the probably clonal pairings (number 20a,b). However this method misses three of the probably clonal pairings, and includes one case that was not identified as clonal using the significance testing strategy. Notice that tumor 20c would have clustered with tumors 20a and 20b if tumors 12a and 01b were not included in the data set, whereas all three pairwise comparisons for this patient are significant with respect to the reference distribution at the 5% level. This shows a drawback of hierarchial clustering in that the diagnosis of any individual patient can be strongly dependent on results from other individual patients.

The summary results for all the similarity measures are presented in Table 2. There are eight analytic strategies: four similarity measures each employed using either hierarchical clustering or significance testing. The table shows the relative numbers of independent and clonal classifications, the proportion of the four definitively clonal pairs that are identified correctly, and the proportion identified as clonal from the larger group of eight pairings that are probably clonal. The methods show substantial variation in their classifications. Although
all methods other than raw correlation detect all or most of the definitely or probably clonal pairs using the significance testing approach, there is an increased tendency to miss these using hierarchical clustering, as logic would predict. Interestingly, simple correlation of the segmented means seems to work well at detecting the clonal pairs, but this method classifies an additional three cases as clonal. Although these cases had no matching point mutations, we cannot be sure that they are independent tumors, i.e., false positives. Readers are encouraged to examine the segmented plots of these and other cases in the supplementary material to our previously published analyses of these data, in which these three cases are numbered 3, 10, and 19 (49). The plots are available in Supplementary Data S1.

Discussion

Analyses of genomic data are challenging because of the sheer volume of data and the complexity of the data summarization tasks required to reduce the data to manageable form. Testing pairs of tumors from the same patient for clonal similarity is especially challenging because it is not obvious how best to summarize the data to extract the signal that is likely to be optimally sensitive to clonal allelic changes. Perhaps as a result, this field is characterized by a multitude of proposed statistical methods. Many of these are adaptations of hierarchical clustering, an exploratory tool with a long history in the field of statistics, but one which has achieved a renaissance of interest in the genomic era. Unfortunately, although this method provides a natural framework for identifying clusters of similar patients from among a larger sample of patients with genomic data, we have shown that it is not a conceptual analytic structure that is suited to clonality testing.

A more appropriate conceptual structure for this kind of analysis is the hypothesis or diagnostic testing framework. In this framework we can assess the plausibility of the tumors being independent by determining whether the chosen similarity measure is representative of the distribution of the measure among independent pairs of tumors, because we can determine this reference distribution by studying the alignments of tumors from different patients. At the outset this approach can only distinguish tumor pairs that are not representative of independent tumor pairs. It does not tell us if the measure of similarity for the tumor pair is representative of tumor pairs that are known to be clonal. The problem is that to obtain a corresponding reference distribution for the alternative clonal hypothesis we would need to assemble a large group of patients from the clinical scenario under investigation whose tumors are known to be clonal. This is problematic because we cannot be certain that any tumor pair in our sample is truly of clonal origin, however similar the array copy number patterns might look. This is actually a familiar conundrum in medical diagnostic testing, the problem of the absence of a gold standard diagnosis (50, 51). It might be tempting in this setting to pick out pairs of tumors whose copy number patterns are sufficiently similar that we can be convinced, regardless of the evidence from conventional pathology, that the tumors are clonal, and construct the reference distribution using only these cases. The problem with this strategy is that it would not be representative of the true reference distribution in that it would have a bias toward clear-cut cases and against equivocal cases, and this would exaggerate the true accuracy of the diagnostic system.

The creation of a suitable measure of similarity for pairs of array data being compared is a pivotal aspect of the analysis. Arrays typically indicate many allelic gains and losses, only some of which may be clonal, the others representing independently occurring allelic changes after the two tumors began to develop independently. It is consequently difficult to figure out how much evidential weight to assign to, say, a single very closely matching pair of gains or losses in the presence of many other non-matching allelic changes elsewhere in the genome. Noise

### Table 2. Summary of classification results from the different measures and methods

<table>
<thead>
<tr>
<th>Measure</th>
<th>Hierarchical clustering classifications</th>
<th>Significance testing classifications</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Independent</td>
<td>Clonal</td>
</tr>
<tr>
<td>Correlation (raw)†</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>Correlation (segmented)§</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>Matching Break Points‡</td>
<td>18</td>
<td>6</td>
</tr>
<tr>
<td>LR‡</td>
<td>18</td>
<td>6</td>
</tr>
</tbody>
</table>

*Proportion of four definitely clonal pairs identified.
†Proportion of eight definitely or probably clonal pairs identified.
‡Correlation of raw marker values (after local averaging).
§Correlation of segmented marker values.
‡Weighted count of matching cut-pints identified by segmentation (Bollet et al.; ref. 40).
†LR measure.
in the marker values adds uncertainty to the detection and comparison of allelic changes. Very noisy arrays are especially difficult to analyze. There is a risk that some observed (small) allelic changes may occur in the germ-line, ensuring that they show matches even in independent tumors. Added to this is the problem that there is no gold standard technique for identifying clonal tumor pairs to use as a benchmark for comparing different methods. Indeed these methods are likely to have the greatest clinical application precisely in these clinical settings where pathologists lack confidence in diagnoses on the basis of conventional gross pathological criteria. Although the results of our simulations in Table 1 and comparative analyses of real data in Table 2 show clearly the deficiencies of hierarchical clustering, they do not show any clear advantage for any one of the selected similarity measures we have used. Establishing the optimal choice of similarity measure is challenging, and is a future research task.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We are grateful to our colleagues Nicolas Girand and William Pao for consenting to the use of data from our joint study of lung cancer (49).

Grant Support

National Cancer Institute, award CA124504.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 09/02/2009; revised 12/10/2009; accepted 12/10/2009; published OnlineFirst 02/23/2010.

References


Clinical Cancer Research

Testing Clonal Relatedness of Tumors Using Array Comparative Genomic Hybridization: A Statistical Challenge

Irina Ostrovnaya and Colin B. Begg


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-09-2398

Cited articles
This article cites 50 articles, 13 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/16/5/1358.full.html#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
/content/16/5/1358.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.