Published in final edited form as:

Biometrics. 2007 June ; 63(2): 522–530. doi:10.1111/j.1541-0420.2006.00681.x.

Statistical Tests for Clonality

Colin B. Begg^{*}, Kevin H. Eng, and Amanda J. Hummer

Memorial Sloan-Kettering Cancer Center, Department of Epidemiology and Biostatistics, 307 E 63rd Street, Room 303, New York, NY 10021

SUMMARY

Cancer investigators frequently conduct studies to examine tumor samples from pairs of apparently independent primary tumors with a view to determining if they share a “clonal” origin. The genetic fingerprints of the tumors are compared using a panel of markers, often representing loss of heterogeneity (LOH) at distinct genetic loci. In this article we evaluate candidate significance tests for this purpose. The relevant information derives from the observed correlation of the tumors with respect to the occurrence of LOH at individual loci, a phenomenon that can be evaluated using Fisher’s Exact Test. Information is also available from the extent to which losses at the same locus occur on the same parental allele. Data from these combined sources of information can be evaluated using a simple adaptation of Fisher’s Exact Test. The test statistic is the total number of loci at which concordant mutations occur on the same parental allele, with higher values providing more evidence in favor of a clonal origin for the two tumors. The test is shown to have high power for detecting clonality for plausible models of the alternative (clonal) hypothesis, and for reasonable numbers of informative loci, preferably located on distinct chromosomal arms. The method is illustrated using studies to identify clonality in contralateral breast cancer. Interpretation of the results of these tests requires caution due to simplifying assumptions regarding the possible variability in mutation probabilities between loci, and possible imbalances in the mutation probabilities between parental alleles. Nonetheless, we conclude that the method represents a simple, powerful strategy for distinguishing independent tumors from those of clonal origin.

Keywords

Clonality; Permutation test; Second primary cancers

1. Introduction

Patients with cancer frequently experience a second malignancy. With the aging of the population, second primary malignancies now account for 16% of all cancer incidences in the USA (Travis et al., 2005). Often these cancers occur in the same organ. For example, patients with breast cancer can experience a new malignancy in the opposite (contralateral) breast, or in the same (ipsilateral) breast if breast tissue remains after the initial surgery. In these circumstances it is important for clinical reasons to know whether the new malignancy is an independent occurrence of cancer, or merely metastatic spread of the original tumor. This distinction is important in determining whether the patient might benefit from systemic chemotherapy (Goldstein et al., 2005). Various pathologic criteria have been used to distinguish second primaries from metastases for reporting to cancer registries (see Schottenfeld, 1996), and for clinical purposes, with histologic type, stage and anatomic

^{*}beggc@mskcc.org.

location being the predominant criteria for distinguishing the new lesions (Noguchi et al., 1994). However, these criteria are arbitrary, and may be subject to considerable error.

In recent years investigators have studied this issue in various organ systems using molecular profiling of cells from pairs of tumors from individual patients, and a large literature of these studies has developed, most prominently in the area of head and neck cancer (Ha and Califano, 2003) and bladder cancer (Hafner et al., 2002). Typically, this involves examination of the tumors for somatic mutations in genes that are frequently altered in cancers of the type under investigation, or loss of heterozygosity (LOH) at mutational hot spots where LOH occurs frequently. LOH is the phenomenon where a marker is heterozygous in normal tissue but has lost an allele in the tumor. Somatic losses (and gains) of alleles occur frequently in tumors, and the pattern of these losses characterizes the events that led to the development of the tumor, and the resulting genetic instability. Thus, by comparing normal tissue and tumor tissue at candidate genetic markers one can determine regions in which LOH has occurred. By establishing the presence or absence of LOH at each of these candidate loci for the two tumors, one can compare the patterns of mutations to see if they appear to be closely matched. If so, the tumors are considered to be “clonal”, that is deriving from a single cell that experienced the pivotal mutations prior to seeding both tumors. Our goal in this article is to develop a formal statistical test to determine whether these patterns are consistent with the hypothesis that the two tumors are independent, or whether the patterns are sufficiently similar to indicate a clonal origin.

The data for an individual patient with two potentially “independent” cancers consist of the presence or absence of LOH at each of a number of candidate genetic loci. If losses occur on both tumors at a given locus one can also determine if the loss has occurred on the same parental allele. This is necessarily the case for clonal mutations. Under our null hypothesis of independence all of these somatic mutations are assumed to have occurred independently.

Clonality represents the phenomenon whereby both tumors are derived from the same cell that experienced the pivotal mutations that led to the cancer. Thus in tumors of clonal origin, initially the mutation patterns in the tumors must have been identical, having occurred in the same cell. Consequently, losses occurring in this stage will be evident on the same parental allele on both tumors. Later as the two clones grow independently, new independent mutations may occur. Thus the mutational patterns that we observe involve, under the alternative, clonal hypothesis, correlation in the outcomes, denuded by the “noise” introduced by late, random somatic mutations.

There is no consistent approach to the statistical analysis of these clonality studies. Investigators frequently classify the pairs of tumors in their study as independent or clonal by informal evaluation of the side by side mutational profiles, sometimes in concert with gross pathological or clinical information. In a review of clonality studies of bilateral breast cancer that we performed as background data for this article, the issue of statistical analysis is frequently not discussed. Occasionally investigators have adopted statistical models constructed specifically for this problem. In two recent studies of ovarian cancer, two groups of investigators approached the issue independently using likelihood ratios to distinguish the two hypotheses (Sieben et al., 2003; Brinkmann et al., 2004). However, these likelihood ratio approaches necessitate the specification of parameters that are unknown at the outset (though they could, in principle, be estimated) such as the marginal mutation frequencies, and the underlying degree of correlation in mutational profiles one would expect in clonal tumors.

Our goal in this article is to formulate a basic statistical test for this problem that allows investigators to test pairs of tumors in individual patients for clonality without the need for

extraneous information. An obvious basic strategy is simply to examine the degree to which the occurrences of LOH between loci are correlated, using a simple 2×2 table. However, the occurrences of allelic losses on the same parental allele at loci with losses on both tumors also provides important information. To accommodate this, we derive two permutation tests, analogous to Fisher's Exact Test, that condition on the observed (marginal) counts of mutational events in each tumor. We compare the operating characteristics of these tests and Fisher's Test. The tests have contrasting properties, and we discuss the circumstances under which one approach might be preferred over the others. To shed further light on the methods we reanalyze data from several articles addressing the hypothesis of clonality between first and second primary breast cancers.

2. Methods

The preponderance of the studies that we analyze in Section 3 used LOH at a set of candidate loci for mutational profiling. For expository convenience in the following we use the terminology of LOH in our description of the statistical methods, and we refer to a specific motivating article as the background (Imyanitov et al., 2002). However, the methods are potentially applicable if the information at a locus is the presence of any somatic mutation, not necessarily LOH. Sometimes the studies can involve individual mutations within a targeted gene that frequently experiences somatic mutations in tumors (such as p53), or allelic gains, alone or in addition to allelic losses. The raw data are displayed in detail in the article by Imyanitov et al., and readers are encouraged to use their Figure 1 as background. Data from 3 selected cases are displayed in our Table 1, using the same nomenclature. In this study 14 genetic loci were tested. At each locus on distinct chromosomal arms several markers were examined to find a heterozygous marker. If none could be found the locus was classified as uninformative for the patient (denoted by “—” in the table). For each informative locus, the marker was examined in each of the two tumors for loss of heterozygosity (LOH). LOH could be present in both, either one or neither tumor (losses are denoted by $\blacktriangle$ in the table, while no LOH is denoted by the symbol “o”). If LOH was present in both tumors, it is further necessary to know if the loss was present on the same parental allele (as would be the case if the mutation was clonal) or different alleles (in which case these mutations are necessarily independent). The latter are denoted using discordant symbols “ $\blacktriangle \blacktriangledown$ ” in the table. Thus for a given patient the informative loci can be classified into distinctly informative groups, with frequencies defined as follows:

a = number of loci with LOH on the same allele in both tumors (concordant mutations),

e = number of loci with LOH in both tumors, regardless of allelic concordance,

f = number of loci with LOH on tumor 1 only (discordant),

g = number of loci with LOH on tumor 2 only (discordant),

h = number of loci with no LOH (concordant normals).

We define $J = e + f + g + h$ as the total number of informative loci. Among these J loci, pairs of tumors that share larger numbers of “concordant” loci are more reflective of clonal tumors, while larger numbers of discordant loci suggest independent origin of the tumors. Note that the $(e - a)$ loci where both tumors have mutations, but where these are on opposite alleles, are considered discordant.

In all that follows we assume that the occurrence of LOH at different markers in the same tumor are independent. This assumption is appropriate for the data from Imyanitov et al., since the 14 loci occur on distinct chromosomal arms. If distinct markers on the same

chromosomal arm are used separately, as some investigators in this field have done, dependencies between these markers are not unlikely since it is common for allelic losses to occur across a broad region of a chromosomal arm or indeed the entire arm. For this reason, we recommend that our methods be used solely for markers on distinct chromosomal arms.

Typical data configurations are set out in Table 1 using data from Imyanitov et al. (2002). Case #3 has 13 informative loci, all of which are concordant, with 4 of these representing concordant mutations on both tumors. At first glance, these two tumors would appear to be clearly clonal, but one certainly needs a statistical test to assess the probability that this configuration could occur in two tumors of independent origin simply by chance. Case #32 also has clonal features in that in 5 of the 13 informative loci there are concordant losses. However, at one locus (22q) the markers are discordant, and there are 3 loci where mutations occur in one tumor but not in the other. Clearly this is a more difficult case to judge informally. Case #22 appears to represent an independent pattern of mutations.

The test statistics considered in the next section are all based on the null distribution of these frequencies under the assumption that the various somatic mutations have occurred independently in the two tumors. This distribution can be viewed as a straightforward compound sampling setup, where in an “outer” 2×2 contingency table the frequencies e, f, g and h are distributed randomly subject to an underlying mutation probability p , assumed common to both tumors. To eliminate dependence on the unknown p , one can condition on the marginal frequencies $m_1 = e + f$ and $m_2 = e + g$, as in Fisher’s Exact Test, in which case

$$\Pr(e|m_1, m_2) = \frac{\binom{m_1}{e} \binom{J-m_1}{m_2-e}}{\binom{J}{m_2}}$$

In the presence of clonality, many of the mutations will be common to both tumors, and so the frequencies will be correlated. Following Cox (1970) we characterize this correlation by the odds ratio, φ , and under the alternative (clonal) hypothesis

$$\Pr(e|m_1, m_2, \varphi) = K^{-1} \binom{m_1}{e} \binom{J-m_1}{m_2-e} \varphi^e, \quad (1)$$

where

$$K = \sum_{j=\max(0, m_1+m_2-J)}^{m_1} \binom{m_1}{j} \binom{J-m_1}{m_2-j} \varphi^j.$$

Pairs of tumors that share a clonal origin will possess at least some mutations that are identical, and possibly others that occurred subsequent to the seeding of the two tumors by the original clone. Such tumors will be characterized by a positive correlation, and so Fisher’s Test is a viable strategy for testing for clonality. However, there is additional information in the loci at which LOH is present in both tumors. Those mutations that occurred in the clonal phase (i.e. in a single cell) are necessarily concordant (on the same allele), while non-clonal mutations occur independently of each other. If we assume that

each parental allele is equally likely to experience the mutation (an assumption that will be discussed in detail in Section 3.3), then the null probability that for any individual locus the

mutations occur on the same allele is $1/2$, and so $\Pr(a|e) = \binom{e}{a} 2^{-e}$. Under the clonal

alternative, $E(a) > \frac{1}{2}E(e)$. We characterize the induced preference for a concordant mutation by the parameter $\pi = E(a)/E(e)$. Later, in Section 3.1 we interpret the parameters φ and π in a more conceptually understandable context. All the other relevant frequencies are determined from a and e , given the fixed margins m_1 and m_2 . [For notational convenience in all that follows we adopt the convention that $m_1 < m_2$, *w.l.o.g.*]

2.1 Candidate Tests

We offer three candidate tests, with an explanation of the rationale for each.

Test 1: Concordant Mutations Test—Here we consider the test statistic to be the total number of concordant mutations, a , and the test is conditional on the outer margins, m_1 and m_2 . The test is significant (at the α level) if $a \geq a_\alpha$, where a_α is the smallest integer such that

$$\sum_{j=a_\alpha}^{m_1} \sum_{i=a_\alpha}^j 2^{-j} \frac{\binom{j}{i} \binom{m_1}{j} \binom{J-m_1}{m_2-j}}{\binom{J}{m_2}} \leq \alpha \quad (2)$$

For a given patient, the p-value can be calculated from the left hand side of (2) by replacing a_α with the observed value of a . The justification for this test is that concordant mutations provide the evidence for clonality, and so we seek to determine whether the number of concordant mutations exceeds the number that would be expected on the basis of chance. The null hypothesis is that $\varphi = 1$ and $\pi = 0.5$.

Test 2: Concordant Loci Test—Although concordant mutations would appear to provide the principal evidence for clonality, the frequent presence of loci in which no mutation occurs in either tumor is also (indirectly) evidence in favor of clonality. Or, to put it differently, a higher value of h necessarily corresponds with lower values of f and g , and f and g represent discordant events that provide the evidence against the clonal (alternative) hypothesis. Thus, the test statistic $s = a + h$, the total of all concordant loci, would appear to be a viable candidate. The null distribution of s is given by

$$\Pr(s|m_1, m_2) = \sum_{k=\max\{0, \text{int}\{(s+1+J-m_1-m_2)/2\}\}}^{\min(s, J-m_2)} 2^{-(m_1+m_2+k-J)} \frac{\binom{m_1+m_2+k-J}{s-k} \binom{m_1}{m_1+m_2+k-J} \binom{J-m_1}{J-m_1-k}}{\binom{J}{m_2}}$$

where $\text{int}\{\cdot\}$ represents the smallest integer greater than or equal to the argument. This test is significant at the α level if $s \geq s_\alpha$, where s_α is the smallest integer such that

$$\sum_{l=s_\alpha}^{J+m_1-m_2} \sum_{k=\max\{0, \text{int}\{(l+1+J-m_1-m_2)/2\}\}}^{\min(l, J-m_2)} 2^{-(m_1+m_2+k-J)} \frac{\binom{m_1+m_2+k-J}{l-k} \binom{m_1}{m_1+m_2+k-J} \binom{J-m_1}{J-m_1-k}}{\binom{J}{m_2}} \leq \alpha \quad (3)$$

The p-value can be obtained by replacing s_α with s in (3). This is a test of the null hypothesis that $\varphi = 1$ and $\pi = 0.5$.

Test 3: Fisher's Exact Test—As a third option we simply consider the strategy of employing Fisher's Exact Test on the “outer” 2×2 table comprising the frequencies e, f, g and h . In this approach we ignore the frequency with which allelic losses on both tumors involve the same parental allele. Although this approach discards information of considerable relevance, it avoids a potentially serious validity problem that will be discussed later in Section 3.3. This is a test of the null hypothesis that $\varphi = 1$.

2.2 Derivation of Statistical Power

In our statistical formulation, we have characterized the correlation in outcomes under the alternative “clonal” hypothesis using two parameters, φ and π . Along with the overall marginal probability of a mutation, denoted p , these parameters allow us to characterize the power of the tests in closed-form. As indicated in (1), the distribution of e , the number of loci with mutations in both tumors, conditional on the marginal totals, m_1 and m_2 , can be expressed as a function of φ . Likewise, within this group of loci with mutations in both tumors, i.e. conditional on e , the distribution of a , the number of concordant mutations, is a function solely of π , the probability that losses in each tumor at the same locus will occur on the same parental allele. In order to calculate power one must average these conditional powers with respect to the distribution of the marginal totals (m_1, m_2) under the alternative hypothesis, and this depends on the overall mutation frequency, p , in addition to φ . Let $\bar{e} = E(e) = J[1 + 2p(\varphi - 1) - \{1 + 4p(1 - p)(\varphi - 1)\}^{1/2}/2(\varphi - 1)]$. Then one can construct the joint probability of (m_1, m_2) by recognizing that the conditional probability of a mutation in

one tumor given a mutation in the other is $\frac{\bar{e}}{p}$, and the conditional probability of a mutation in the tumor given no mutation in the other is $\frac{p - \bar{e}}{1 - p}$. Consequently

$$\Pr(m_1, m_2) = 2^{\delta(m_1, m_2)} \sum_{i=\max(0, m_1+m_2-J)}^{m_1} \binom{J}{m_1} \binom{m_1}{i} \binom{J-m_1}{m_2-i} \bar{e}^{-i} (p - \bar{e})^{m_1+m_2-2i} (1 + \bar{e} - 2p)^{J-m_1-m_2+i},$$

where $\delta(m_1, m_2) = 1$ if $m_1 \neq m_2$ and 0 otherwise. The power for each test can be calculated as follows. For the Concordant Mutations Test

$$\text{Power} = \sum_{m_1=0}^J \sum_{m_2=m_1}^J \Pr(m_1, m_2) \sum_{j=a_\alpha}^{m_1} \sum_{i=a_\alpha}^j K^{-1} \binom{j}{i} \binom{m_1}{j} \binom{J-m_1}{m_2-j} \varphi^j \pi^i (1 - \pi)^{j-i},$$

where K is as defined in (1). Note that in the preceding formula a_α is a function of (m_1, m_2), as defined in (2). For the Concordant Loci Test

$$Power = \sum_{m_1=0}^J \sum_{m_2=m_1}^J \Pr(m_1, m_2)$$

$$\sum_{l=s_\alpha}^{J+m_1-m_2} \sum_{k=\max\{0, \text{int}\{(l+1+J-m_1-m_2)/2\}\}}^{\min(l, J-m_2)} K^{-1} \binom{m_1+m_2+k-J}{l-k} \binom{m_1}{m_1+m_2+k-J} \binom{J-m_1}{J-m_1-k} (\varphi(1-\pi))^{m_1+m_2+k-J} \left(\frac{\pi}{(1-\pi)}\right)^{(1-k)}.$$

For Fisher's Test the power is calculated using

$$Power = \sum_{m_1=0}^J \sum_{m_2=m_1}^J \Pr(m_1, m_2) \sum_{e=e_\alpha^*}^{m_1} \Pr(e|m_1, m_2, \varphi),$$

where $\Pr(e | m_1, m_2, \varphi)$ is as defined in (1) and e_α^* is the α -level critical value of the test for marginal totals m_1 and m_2 .

For the purposes of comparing the properties of the tests in the next Section, in addition to calculating power as indicated above we have also calculated “calibrated” powers, where a calibrated power is the power adjusted to a size of exactly α . This eliminates anomalies in the comparisons that are solely due to the discreteness of the tests, and the resulting conservative sizes. Consider the Concordant Mutations Test. Let

$$t(a, \varphi, \pi) = \sum_{m_1=0}^J \sum_{m_2=m_1}^J \Pr(m_1, m_2) \sum_{j=a}^{m_1} \sum_{i=a}^j K^{-1} \binom{j}{i} \binom{m_1}{j} \binom{J-m_1}{m_2-j} \varphi^i \pi^i (1-\pi)^{j-i}.$$

Then the size of the test is $t(a_\alpha, 1, 0.5)$ and the true power is $t(a_\alpha, \varphi, \pi)$. Setting $r = \{\alpha - t(a_\alpha, 1, 0.5)\} / \{t(a_\alpha - 1, 1, 0.5) - t(a_\alpha, 1, 0.5)\}$ the calibrated power of the test, corresponding to a calibrated size of α , is given by $t(a_\alpha, \varphi, \pi) + r\{t(a_\alpha - 1, \varphi, \pi) - t(a_\alpha, \varphi, \pi)\}$. An analogous approach is used to calibrate the other tests.

3. Results

In this section we evaluate the size and power of the tests under various parametric configurations, we analyze some available data sets from the literature, and we explore the validity of the tests under various conceptual scenarios.

3.1 Operating Characteristics of the Tests

The powers of these tests depend on the number of loci investigated, J , and the probability of a mutation at any given locus, p . However, they also depend critically on the strength of the correlations, represented by φ and π , that we might expect when two tumors are clonal. Consider the following model for the generation of the degree of correlation. The marginal probability of a mutation, p , is partitioned into q_b and q_a where q_b represents the somatic mutations that occur in the initiating clonal cell, while q_a represents the “random” mutations that occur “after” the cell separates into two distinct clones. This conceptual model has been adopted by both Sieben et al. (2003) and Brinkmann et al. (2004) in their strategies based on

the likelihood ratio. In this framework, identical (concordant) mutations are distributed to both tumors with probability q_b , while subsequent random mutations, which could be concordant by chance, are distributed to each tumor with probability $q_a = p - q_b$. Under this

model $E(e) = J[q_b + q_c^2(1 - q_b)]$ and $E(a) = J[q_b + \frac{1}{2}q_c^2(1 - q_b)]$, where $q_c = q_a/(1 - q_b)$, and it follows that

$$\varphi = \frac{q_b + q_c^2(1 - q_b)}{q_c^2(1 - q_b)}.$$

and

$$\pi = \left\lceil \frac{q_b + \frac{1}{2}q_c^2(1 - q_b)}{q_b + q_c^2(1 - q_b)} \right\rceil.$$

In Table 2, the characteristics of the tests are displayed for various configurations under this structure, based on a significance level of 0.05. The calibrated values are displayed in parentheses. Note that configurations in which $q_b = 0.0$ correspond to the null hypothesis of independence.

The major observations from the table are as follows. First, the Concordant Mutations Test and the Concordant Loci Test possess very similar power. The overall power for these latter two tests is high if we can be confident that the “signal” is strong. That is, when the bulk of the mutations occur in the clonal phase of development, $q_b \gg q_a$, we can obtain well in excess of 90% power in a study with as few as 10 loci when $p = 0.5$, with a corresponding power of about 80% when $p = 0.3$. When the marginal probability of a mutation is small ($p = 0.1$) we need to examine at least 20 loci to achieve good power. The tests are notably less powerful when about half of the mutations occur in the clonal phase ($q_b = q_a$), and in these cases one would need about 30 loci with relatively high mutation probabilities to achieve good power. Fisher’s Test, which does not use the information about allelic concordances, is substantially less powerful, but it is nonetheless capable of achieving high power if the signal is strong, the mutation probability is relatively high, and the number of loci examined is at least 20.

3.2 Data Examples

We have assembled several examples from the literature, all of which are addressed at determining the clonality of breast tumors. We report here on those studies where the authors published the data in sufficient detail for us to construct the tests for each patient. The study by Imyanitov et al (2002) is a good example, since there is a broad range of subjects (28 cases) and the number of loci examined (14) is typical for this type of study. We previously highlighted three representative patients from this study, one of which appeared clonal on an informal basis (Case #3), one of which appears to be random (Case #22), and one (Case # 32) which has evidence for and against clonality. The p-values from our tests for these patients are set out in Table 1. The results show that for all three tests the clonality of Case #3 is confirmed with high statistical significance, and the apparent independence in Case #22 is also suggested by the absence of significance. The two tests that take advantage of the information on allelic concordance demonstrate significant

evidence of clonality for the “ambiguous” case (Case #32) while Fisher’s Test is not significant at $p=0.09$.

When the tests are applied to all the patients in the data set, we obtain the following classifications. The Concordant Mutations Test classifies 7 of the 28 patients as significant for clonality, the Concordant Loci Test classifies 9 as significant, and Fisher’s Test classifies 3 as significant. It is of note that in their own interpretation of these data, Imyanitov et al. conclude that Case #3 is not clonal based on clinical criteria, specifically the fact that the tumor in the left breast contained an in situ component and the right tumor contained a higher degree of differentiation. The conclusiveness of judgments such as this regarding clonality is open to question. Interestingly, Imyanitov et al. concluded that there was unambiguous evidence for clonality for only one of their 28 cases (Case #10), in that the clinical parameters did not conflict with the molecular evidence favoring clonality. In this case, 5 loci had concordant mutations, and only one other mutation was observed, out of 14 informative loci. This case is highly statistically significant for all three tests.

Chunder et al (2004) examined four patients with either bilateral or multifocal breast cancer using LOH at a total of 26 markers on 10 distinct chromosomal arms. The number of informative loci ranges from 16 to 24. Three of the cases showed random patterns ($p = 0.58$, $p = 0.89$ and $p = 1.00$ for the Concordant Mutations Test), but one of the cases had concordant mutations at 9 loci and no discordances among 24 informative loci, and is highly significant for clonality on all three tests. Note that in applying the tests to these data we have had to assume that the losses on the same chromosomal arms are independent.

Kollias et al. (2000) studied bilateral breast cancer in 31 patients, but they used only 6 markers, one of which was a control marker that registered a loss in only one patient. None of the tests were significant for any of the patients, and it is certainly possible that this was the result of low power, though none of the 31 patients experienced identical mutations patterns in the two tumors.

Regitnig et al. (2004) studied 8 patients with bilateral breast cancer, five of whom also had tissue samples from local recurrences, using 13 loci on 10 distinct chromosomal arms (informative markers ranged from 5 to 8). Very few concordant mutations were observed in the pairs of primaries and none of the tests were significant.

Finally, Goldstein et al. (2005) examined the clonality of new ipsilateral breast cancers with reference to their primary tumors (two cancers in the same breast) in 26 patients using 20 markers on 14 distinct chromosomal arms (10 to 17 informative). Their patients split into two distinct groups: those where markers on both tumors were predominately concordant, and those where the markers were predominately discordant. Although the data were not reported in sufficient detail for us to determine the marginal totals, analyses of all the possible configurations of the marginal totals shows that in the former group of 18 patients the Concordant Mutations Test and the Concordant Loci Test were significant in at least 16 and possibly 18 of the patients, and at worst marginally significant in the two doubtful cases. We present this example to demonstrate the high power of these tests. As shown in Section 2.2, the power is dependent on the “signal” under the alternative. However, ipsilateral breast cancer has traditionally been assumed to be clonal in most patients, and so the fact that significance of the test is attainable in most presumptively clonal patients in a study with only 10 to 17 informative markers demonstrates that in this setting relatively small sample sizes (i.e. numbers of loci examined) are needed to rule in or rule out clonality with acceptable precision.

3.3 Validity Issues

The calculations in Table 2 demonstrate that the tests are valid for testing the null hypothesis of independence when the mutations in the two tumors are generated independently ($\phi = 1$), the mutations are allocated to the two paternal alleles with equal probability ($\pi = 0.5$), and where the individual events are identically distributed with a common probability of mutations, p , for both tumors and at each locus. However, there are two issues that have the potential to seriously compromise the validity of the tests in real applications of this methodology. In practice, each locus will have a different propensity for experiencing a mutation. For example among the 14 loci in the data of Imyanitov et al., the empirical mutation rates vary from a low of 28% (8/28) on chromosome 5q to a high of 73% (41/56) for chromosome 17q. Recognizing that these empirical rates will be more variable than the underlying true rates, they nonetheless demonstrate considerable variation. Adopting a random effects model, the moment estimator of the variance between loci (Bohning et al., 2002) is estimated to be 0.023. In Table 3 we have evaluated the sizes of the tests in the setting where the loci are assigned to three groups of approximately equal frequency, but with differing mutation probabilities and with a variance between loci that is of the same magnitude as the preceding value of 0.023 estimated from the Imyanitov data, and also from hypothetical configurations with double this variance estimate. It is clear from the results that heterogeneity of this nature does inflate the test size to some extent. For the Concordant Mutations test, the inflation in test size is modest, and indeed the true uncalibrated size does not exceed the nominal value of 0.05 for any of the configurations corresponding to the estimated variance of 0.023, although the size is inflated above the nominal level for the configurations with the much larger variance. The anti-conservative bias is considerably more pronounced for the Concordant Loci Test and for Fisher's Exact Test.

A different validity concern is that at individual loci allelic losses are not equally probable on the two parental alleles. If loss of heterozygosity is indeed relevant to carcinogenesis at a given locus, then it seems likely that allelic loss of the more important allele will occur in tumors with probability greater than $1/2$. Let us assume that at a given locus the more important allele is lost with probability $r > 1/2$. Then even under the independence model the probability that losses on both tumors will be concordant is $\pi = r^2 + (1 - r)^2 > 1/2$. The data from the cited studies in Section 3.2 are very inconsistent on this issue. In the studies by Regitnig et al. and Chunder et al., if we exclude the single case that is clearly clonal, then among loci with LOH on each tumor concordant and discordant mutations are relatively evenly distributed, 11 versus 16 respectively. Conversely in the studies by Imyanitov et al. and Kollias et al. there are strong preponderances of concordant mutations. Indeed in the study by Kollias et al. all 19 such loci were concordant, a fact that was commented on by the authors. Curiously, in the study of ipsilateral breast cancers by Goldstein et al. the patients seem to split into two distinct groups, those classified as clonal versus those classified as independent by the authors. In the "clonal" group 70 of 72 paired loci were concordant, as one might expect, but curiously in the "independent" group the vast majority of pairs were discordant (24 versus 4). The bottom line is that these data are mutually inconsistent, but we must be concerned that the concordance probability will be elevated above 0.5 in practice to an unknown degree.

In Table 4 the sizes of the tests are evaluated with data generated independently for the two tumors, but with an elevated concordance probability of 0.6 or 0.7. The Concordant Mutations Test is the one most affected, with serious anti-conservative bias when the mutation probability approaches 0.5 and the concordance probability approaches 0.7. Fisher's Test is unaffected by this bias since it does not utilize the information on allelic concordance.

4. Discussion

Although many of the studies of clonality in the literature have been published without accompanying statistical analysis, but with simple judgments about the apparent concordance in the genetic fingerprints of the tumors, some investigators have suggested alternative methods. One method that has seen considerable application is the use of a statistical “measure of clonality”, originally proposed by Kuukasjarvi et al. (1997), and adopted by, among others, Jiang et al. (2005). This measure involves solely the concordant mutations that have been observed, and the measure is calculated as $1 - \prod p_i^2$, where p_i is the mutation probability at the i^{th} concordant locus. Since this method does not consider the loci examined which are not concordant, the sampling frame is not properly embedded in the method. That is, the method will give the same evidence for clonality when, say, three loci are concordant, regardless of the number of loci that are examined in the first place. Sieben et al. (2003) suggest a likelihood ratio approach. They too consider the overall mutation probability as composed of pre-clonal and post-clonal components, and they construct a likelihood ratio statistic under the assumption that these two probabilities are equal and known. Their method was used by Goldstein et al. (2005) in their study of clonality of ipsilateral breast cancer recurrences. Brinkmann et al. (2004) also employ a likelihood ratio approach in which (under the clonal hypothesis) 10% of the genetic events are assumed to occur after the tumors diverge, with the remaining 90% of events being clonal events. The major disadvantage of these approaches is that they rely on an alternative hypothesis for which these parametric values must be pre-specified, with no good information available about their likely values.

Our approach seeks to distinguish independent tumors from those of clonal origin on the basis of a simple statistical test that is entirely self-contained to the individual patient and does not rely on specifying any unknown parametric values. The test involves solely concordance data from the two tumors under comparison, and does not rely on being a part of a larger data set. This is quite important since different studies use different marker sets, and even for an approach using a standardized set of markers the set of informative markers for an individual patient necessarily varies from patient to patient. Also, it is entirely possible that the overall mutation probability may vary greatly from patient to patient, either because of host susceptibility to allelic losses, or merely because the tumors in different patients may be observed at different stages of genetic instability. Indeed, the observed mutation frequency in the Imyanitov et al. data varies between patients from a low of 15% to a high of 88%. Consequently, a patient-specific analytic strategy is appealing, and the traditional strategy of conditioning on the margins, m_1 and m_2 , is a natural way to account for the fact that patients may vary in their overall mutation probability.

A straightforward application of Fisher’s Exact Test in this setting can deliver quite high power, provided that the signal is strong, i.e. the preponderance of somatic mutations occur in the original clone rather than in the tumors’ subsequent independent growth phases, and that the overall mutation frequency is relatively high, 30% or greater (Table 2). However, there is clearly substantial extra information in the relative frequency of concordant mutations at those loci experiencing mutations in both tumors. Our two new proposed tests capture this additional information with similar increases in power (Table 2). However, all three tests must be interpreted with caution. Two factors are likely to lead to anti-conservative p-values when the tumors are actually of independent origin. The first is variation in the mutation probabilities between loci. Our examination of this issue in Table 3 shows that all three tests are affected by this issue, but that the Concordant Mutations Test is the most robust of the three. The second problem is the assumption that a mutation will occur in the paternal versus the maternal alleles with equal probability. In fact, since these

losses on tumors are more likely to occur in the neighborhood of a tumor suppressor gene (Pinkel et al. 1998), it is quite probable that there will be selection in favor of a loss on the allele that has the greatest impact on the functioning of the adjacent tumor suppressor gene. As a result, there will be a preponderance of concordant versus discordant mutations even when the tumors have arisen independently. Our examination of this phenomenon (in Table 4) shows that both of the new tests can be substantially compromised, while Fisher's Test, which does not use this information, is unaffected.

Based on these results, we believe that the Concordant Mutations Test is the preferred approach. However, one cannot be fully confident of interpreting the p-value as it is conventionally understood. That is, the false positive rate may be considerably higher in practice than is implied by the p-value. As such, the p-value in this context more properly reflects a measure for distinguishing clonal from independent tumors, but with uncertainty about the true error rates. In principle, the error rates might be calibrated by examining them in the context of a discriminant analysis in which a large sample of patients with bilateral tumors are classified as either independent or clonal, and where the reference distribution for the test statistic is evaluated using the patients with "independent" tumors. Indeed Schlecter et al (2004) have approached the problem from such a discriminant analysis perspective. In their method a training set of independent tumors was analyzed in a pair-wise fashion to obtain a distribution of similarity scores for independent tumors. The similarity scores of pairs of tumors from test patients were then compared with this distribution, and outliers were classified as clonal. Unfortunately we have no "gold standard" diagnosis for such a classification, and indeed this article can be viewed as a search for a better gold standard.

The studies that we re-examined varied greatly in the numbers of markers evaluated. This ranged from 5 in the study by Kollias et al. to 26 in the study by Chunder et al. Our results suggest that in studies of this nature one would want to evaluate at least 20 markers to be able to distinguish clonal from independent tumors with high accuracy, especially since a proportion of the markers will typically be uninformative. However, the selection of markers also must be made with caution. Allelic losses frequently encompass large genetic regions, and often the entire chromosome arm. Consequently markers in the same genetic region will necessarily be correlated, and the only way to be confident that the markers are statistically independent is to choose at most one informative marker per chromosome arm, as in the study by Imyanitov et al., thereby greatly limiting the potential sample sizes. If the data come from array-based technologies, such as array comparative genomic hybridization (CGH), then entirely different statistical techniques are needed to take account of these dependencies, and to map out the precise regions of allelic loss or gain (Olshen et al., 2004) in preparation for the statistical comparison of the location of these regions between the two tumors.

In summary, despite the very limited use of data and assumptions, the Concordant Mutations Test appears to be a powerful strategy for distinguishing clonality from independence, provided that a sufficient number of markers are examined for LOH, although a stricter decision criterion than $p < 0.05$ is advisable if one wishes to rule out independence with high confidence.

Acknowledgments

The research was supported by the National Cancer Institute, award number CA098438.

References

- Bohning D, Malzahn U, Dietz E, Schlattmann P, Viwatwongkasem C, Biggeri A. Some general points in estimating heterogeneity variance with the DerSimonian--Laird estimator. *Biostatistics* 2002;3:445–457. [PubMed: 12933591]
- Brinkmann D, Ryan A, Ayhan A, McCluggage WG, Feakins R, Santibanez-Korf MF, Mein CA, Gayther SA, Jacobs IJ. A molecular genetic and statistical approach for the diagnosis of dual-site cancers. *Journal of the National Cancer Institute* 2004;96:1441–6. [PubMed: 15467033]
- Chunder N, Roy A, Roychondhury S, Panda CK. Molecular study of clonality in multifocal and bilateral breast tumors. *Pathology – Research and Practice* 2004;200:735–741.
- Cox, DR. *The Analysis of Binary Data*. Chapman and Hall; London: 1970. p. 48-52.
- Goldstein NS, Vicini FA, Hunter S, Odish E, Forbes S, Kraus D, Kestin L. Molecular clonality determination of ipsilateral recurrence of invasive breast carcinomas after breast-conserving therapy. *American Journal of Clinical Pathology* 2005;123:679–689. [PubMed: 15981807]
- Ha PK, Califano JA. The molecular biology of mucosal field cancerization of the head and neck. *Critical Reviews in Oral Biology and Medicine* 2003;14:363–9. [PubMed: 14530304]
- Hafner C, Kneuchel R, Stoehr R, Hartmann A. Clonality of multifocal urothelial carcinomas: 10 years of molecular genetic studies. *International Journal of Cancer* 2002;101:1–6.
- Imyanitov EN, Susptsin EN, Grigoriev MY, Togo AV, Kuligina ES, Belogubova EV, Pozharisski KM, Turkevich EA, Rodriguez C, Cornelisse CJ, Hanson KP, Theillet C. Concordance of allelic imbalance profiles in synchronous and metachronous bilateral breast carcinomas. *International Journal of Cancer* 2002;100:557–64.
- Jiang JK, Chen YJ, Lin CH, Yu IT, Lin JK. Genetic changes and clonality relationship between primary colorectal cancers and their pulmonary metastases – An analysis by comparative genomic hybridization. *Genes, Chromosomes and Cancer* 2005;43:25–36. [PubMed: 15723340]
- Kollias J, Man S, Marafie K, Carpenter S, Pinder S, Ellis IO, Blamey RW, Cross G, Brook JD. Loss of heterozygosity in bilateral breast cancer. *Breast Cancer Research and Treatment* 2000;64:241–251. [PubMed: 11200774]
- Kuukasjarvi T, Karhu R, Tanner M, Kahkonen M, Schaffer A, Nupponen N, Pennanen S, Kallioniemi A, Kallioniemi OP, Isola J. Genetic heterogeneity and clonal evolution underlying development of asynchronous metastasis in human breast cancer. *Cancer Research* 1997;57:1597–1604. [PubMed: 9108466]
- Noguchi S, Motomura K, Inaji H, Imaoka S, Kayama H. Differentiation of primary and secondary breast cancer with clonal analysis. *Surgery* 1994;115:458–62. [PubMed: 7909387]
- Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004;5:55–72.
- Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG. High resolution analysis of DNA copy number variation using comparative genomic hybridization to micro-arrays. *Nature Genetics* 1998;20:207–11. [PubMed: 9771718]
- Regitnig P, Ploner F, Moderbacher M, Lax SF. Bilateral carcinomas of the breast with local recurrence: analysis of genetic relationship of the tumors. *Modern Pathology* 2004;17:597–602. [PubMed: 15017434]
- Schlechter BL, Yang Q, Larson PS, Golubeva A, Blanchard RA, de las Morenas A, Rosenberg CL. Quantitative DNA fingerprinting may distinguish new primary breast cancer from disease recurrence. *Journal of Clinical Oncology* 2004;22:1830–8. [PubMed: 15143075]
- Schottenfeld, D. Multiple primary cancers. In: Schottenfeld, D.; Fraumeni, JF., editors. *Cancer Epidemiology and Prevention*. 2. Oxford University Press; New York: 1996. p. 1370-90.
- Sieben NLG, Kolkman-Uljee SM, Flanagan AM, le Cessie S, Cleton-Jansen AM, Cornelisse CJ, Fleuren GJ. Molecular genetic evidence for monoclonal origin of bilateral ovarian serous borderline tumors. *American Journal of Pathology* 2003;162:1095–1101. [PubMed: 12651602]
- Travis LB, Rabkin CS, Brown LM, Allen JM, Alter BP, Ambrosone CB, Begg CB, Caporaso N, Chanock S, DeMichele A, Figg WD, Gosporadowitz MK, Hall EJ, Hisada M, Inskip P, Kleinerman R, Little JB, Malkin D, Ng AG, Offit K, Pui CH, Robison LL, Rothman N, Shields

PG, Strong L, Taniguchi T, Tucker MA, Greene MH. Cancer survivorship – genetic susceptibility and second primary cancers: research strategies and recommendations. Journal of the National Cancer Institute 2006;98:15–25. [PubMed: 16391368]

Table 1
Selected examples of individual cases. Data obtained from Imyanitov et al. (2002).

Locus ¹	Case #3		Case #32		Case #22	
	L	R	L	R	L	R
1p	○	○	○	○	○	○
1q	○	○	○	○	○	▲
3p	▲	▲	○	▲	▲	▼
5q	○	○	○	▲	○	○
6q	▲	▲	▲	▲	▲	▲
8p	○	○	▲	▲	—	—
11p	—	—	—	—	—	—
11q	○	○	○	○	○	○
13q	○	○	▲	▲	▲	▼
16q	▲	▲	▲	▲	—	—
17p	○	○	○	○	○	▲
17q	▲	▲	▲	▲	▲	▲
18q	○	○	▲	○	○	▲
22q	○	○	▲	▼	—	—
Frequencies ²						
a	4		5		2	
e	4		6		4	
f	0		1		0	
g	0		2		3	
h	9		4		3	
m ₁	4		7		4	
m ₂	4		8		7	
J	13		13		10	
Tests						
Concordant Mutations	p < 0.001		p = 0.02		p = 0.44	
Concordant Loci	p < 0.001		p = 0.01		p = 0.18	
Fisher's Exact	p = 0.001		p = 0.09		p = 0.17	

¹ Absence of LOH is denoted by $\circ$, LOH is denoted by $\blacktriangle$, and losses on different alleles on the two tumors are denoted $\blacktriangle$ and $\blacktriangledown$.

² The frequencies are as defined in the text.

Table 2

Size¹ and power² of the tests

Mutational Probabilities		Number of Loci (J)	Concordant Mutations Test		Concordant Loci Test		Fisher's Exact Test	
p	q _b		q _a					
0.1	0.00	10	0.10	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.00 (0.05)
0.1	0.00	20	0.10	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)
0.1	0.00	30	0.10	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)
0.1	0.05	10	0.05	0.20 (0.24)	0.20 (0.26)	0.05 (0.03)	0.05 (0.03)	0.05 (0.03)
0.1	0.05	20	0.05	0.42 (0.45)	0.43 (0.48)	0.24 (0.04)	0.24 (0.04)	0.24 (0.04)
0.1	0.05	30	0.05	0.55 (0.59)	0.56 (0.61)	0.41 (0.05)	0.41 (0.05)	0.41 (0.05)
0.1	0.09	10	0.01	0.54 (0.56)	0.54 (0.57)	0.19 (0.00)	0.19 (0.00)	0.19 (0.00)
0.1	0.09	20	0.01	0.82 (0.84)	0.82 (0.84)	0.55 (0.03)	0.55 (0.03)	0.55 (0.03)
0.1	0.09	30	0.01	0.92 (0.93)	0.92 (0.93)	0.86 (0.12)	0.86 (0.12)	0.86 (0.12)
0.3	0.00	10	0.30	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)
0.3	0.00	20	0.30	0.01 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)
0.3	0.00	30	0.30	0.02 (0.05)	0.03 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)
0.3	0.15	10	0.15	0.34 (0.41)	0.32 (0.41)	0.11 (0.10)	0.11 (0.10)	0.11 (0.10)
0.3	0.15	20	0.15	0.64 (0.70)	0.66 (0.71)	0.35 (0.27)	0.35 (0.27)	0.35 (0.27)
0.3	0.15	30	0.15	0.82 (0.86)	0.82 (0.85)	0.56 (0.52)	0.56 (0.52)	0.56 (0.52)
0.3	0.25	10	0.05	0.79 (0.81)	0.76 (0.82)	0.49 (0.09)	0.49 (0.09)	0.49 (0.09)
0.3	0.25	20	0.05	0.97 (0.98)	0.97 (0.98)	0.92 (0.59)	0.92 (0.59)	0.92 (0.59)
0.3	0.25	30	0.05	0.99 (0.99)	0.99 (0.99)	0.99 (0.89)	0.99 (0.89)	0.99 (0.89)
0.5	0.00	10	0.50	0.01 (0.05)	0.02 (0.05)	0.01 (0.05)	0.01 (0.05)	0.01 (0.05)
0.5	0.00	20	0.50	0.02 (0.05)	0.03 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)
0.5	0.00	30	0.50	0.02 (0.05)	0.03 (0.05)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)
0.5	0.25	10	0.25	0.39 (0.47)	0.38 (0.47)	0.09 (0.23)	0.09 (0.23)	0.09 (0.23)
0.5	0.25	20	0.25	0.70 (0.76)	0.69 (0.73)	0.27 (0.58)	0.27 (0.58)	0.27 (0.58)
0.5	0.25	30	0.25	0.87 (0.90)	0.85 (0.87)	0.43 (0.78)	0.43 (0.78)	0.43 (0.78)
0.5	0.45	10	0.05	0.96 (0.97)	0.95 (0.96)	0.69 (0.50)	0.69 (0.50)	0.69 (0.50)
0.5	0.45	20	0.05	0.99 (0.99)	0.99 (0.99)	0.98 (0.97)	0.98 (0.97)	0.98 (0.97)
0.5	0.45	30	0.05	1.00 (1.00)	1.00 (1.00)	1.00 (0.99)	1.00 (0.99)	1.00 (0.99)

¹ A significance level of $\alpha = 0.05$ is used

² The calibrated powers are in parenthesis. The displayed values are calibrated in such a way that the size of each test is 0.05, to facilitate comparison of the power between the tests (see text for details).

Table 3

Test size with variable mutation probabilities[/]

Configurations of Mutational Probabilities			Number of Loci	Concordant Mutations Test	Concordant Loci Test	Fisher's Exact Test
Variance = 0.023						
30% @ 0.01	40% @ 0.1	30% @ 0.36	10	0.02	0.03	0.06
30% @ 0.01	40% @ 0.1	30% @ 0.36	20	0.05	0.09	0.11
30% @ 0.01	40% @ 0.1	30% @ 0.36	30	0.06	0.12	0.16
30% @ 0.11	40% @ 0.30	30% @ .49	10	0.03	0.03	0.02
30% @ 0.11	40% @ 0.30	30% @ .49	20	0.04	0.06	0.05
30% @ 0.11	40% @ 0.30	30% @ .49	30	0.05	0.09	0.08
30% @ 0.31	40% @ 0.50	30% @ .69	10	0.03	0.05	0.07
30% @ 0.31	40% @ 0.50	30% @ .69	20	0.04	0.06	0.10
30% @ 0.31	40% @ 0.50	30% @ .69	30	0.05	0.08	0.12
Variance = 0.046						
30% @ 0.03	40% @ 0.30	30% @ 0.57	10	0.04	0.06	0.04
30% @ 0.03	40% @ 0.30	30% @ 0.57	20	0.07	0.14	0.11
30% @ 0.03	40% @ 0.30	30% @ 0.57	30	0.10	0.20	0.18
30% @ 0.23	40% @ 0.50	30% @ 0.77	10	0.04	0.07	0.03
30% @ 0.23	40% @ 0.50	30% @ 0.77	20	0.06	0.11	0.10
30% @ 0.23	40% @ 0.50	30% @ 0.77	30	0.08	0.15	0.15

[/] A significance level of $\alpha=0.05$ is used.

Table 4

Test size under differential allelic mutation probabilities^J

Mutational Probabilities		Number of Loci	Concordant Mutations Test	Concordant Loci Test	Fisher's Exact Test
p	π				
0.3	0.6	10	0.02	0.02	0.01
0.3	0.6	20	0.04	0.04	0.02
0.3	0.6	30	0.04	0.05	0.02
0.3	0.7	10	0.03	0.03	0.01
0.3	0.7	20	0.06	0.06	0.01
0.3	0.7	30	0.07	0.07	0.02
0.5	0.6	10	0.03	0.04	0.01
0.5	0.6	20	0.05	0.06	0.02
0.5	0.6	30	0.08	0.07	0.02
0.5	0.7	10	0.07	0.06	0.01
0.5	0.7	20	0.11	0.09	0.02
0.5	0.7	30	0.16	0.12	0.03

^J A significance level of $\alpha=0.05$ is used.