Harvard University
Harvard University Biostatistics Working Paper Series

Year 2007 *Paper* 67

Assessment of a CGH-based Genetic Instability

David A. Engler[*] Yiping Shen[†]
J F. Gusella[‡] Rebecca A. Betensky[**]

[*]Harvard University, engler@fas.harvard.edu
[†]Center for Genetic Research, MGH, yshen3@partners.org
[‡]gusella@helix.mgh.harvard.edu
[**]Harvard University, betensky@hsph.harvard.edu

Assessment of aCGH-Based Genetic Instability

David A. Engler
Department of Biostatistics, Harvard University
655 Huntington Avenue, Boston, MA 02115
email: engler@fas.harvard.edu

Yiping Shen
Center for Human Genetic Research, MGH
185 Cambridge Street, Boston, MA 02114

James F. Gusella
Center for Human Genetic Research, MGH
185 Cambridge Street, Boston, MA 02114

Rebecca A. Betensky
Department of Biostatistics, Harvard University
655 Huntington Avenue, Boston, MA 02115
email: betensky@hsph.harvard.edu

ABSTRACT

Array-based Comparative Genomic Hybridization (aCGH) is a microarray-based technology that assists in identification of DNA sequence copy number changes across the genome. Examination of differences in instability phenotype, or pattern of copy number alterations, between cancer subtypes can aid in classification of cancers and lead to better understanding of the underlying cytogenic mechanism. Instability phenotypes are composed of a variety of copy number alteration features including height or magnitude of copy number alteration level, frequency of transition between copy number states such as gain and loss, and total number of altered clones or probes. That is, instability phenotype is multivariate in nature. Current methods of instability phenotype assessment, however, are limited to univariate measures and are therefore limited in both sensitivity and interpretability. In this paper, a novel method of instability assessment is presented that is based on the Engler et al. (2006) pseudolikelhood approach for aCGH data analysis. Through use of a pseudolikelihood ratio test (PLRT), more sensitive assessment of instability phenotype differences between cancer subtypes is possible. Evaluation of the PLRT method is conducted through analysis of a meningioma data set and through simulation studies. Results are shown to be more accurate and more easily interpretable than current measures of instability assessment. Software for this approach is available at
http://www.biostat.harvard.edu/~betensky/papers.html.

1 Introduction

Alteration in DNA sequence copy number has been shown to be associated with cancer development. In an attempt to ascertain copy number alterations in tumor samples, array-based comparative genomic hybridization (aCGH) is used to compare genetic material obtained from tumor and reference samples. Both sets of samples are labeled with fluorescent dyes and are hybridized to an array. The resultant set of fluorescence intensity ratios (on the log_2 scale) can then be analyzed for changes in ratio magnitude. Significant changes in magnitude are presumably indicative of copy number alteration. To date, a number of methods of aCGH analysis have been developed that attempt to identify regions of copy number alteration from sets of log_2 ratios (*e.g.,* Hodgson et al. (2001); Olshen et al. (2004); Hupe et al. (2004); Fridlyand et al. (2004); Wang et al. (2005); Engler et al. (2006)).

The identification of genomic gains and losses has a number of possible uses. The ascertainment of specific genetic areas at which loss or gain are associated with cancer prognosis can lead to improved therapeutic approaches. Location of genes potentially associated with cancer development and progression can result in improved understanding of cancer etiology. Assessment of the DNA sequence copy number loss and gain "phenotype", or pattern of gain and loss, across the genome can be used both to accurately subtype cancers and to understand the underlying cytogenic mechanisms. Shen et al. (2007), for example, examined the differences in genetic instability between different pathological grades of meningioma. Specifically, comparisons between benign, atypical, and malignant tumor groups are of interest.

In the past several years, there has emerged a small body of literature in which various measures of aCGH-based genetic instability have been used. (*e.g.,* Peng et al. (2003); Paris

et al. (2004); Bernardini et al. (2005); Lambros et al. (2005); Blaveri et al. (2005); Fridlyand et al. (2006); Herzog et al. (2006)). Because these are univariate measures, they often do not distinguish between different types of genetic instability. Additionally, many current assessments occur in a second stage analysis following methods of copy number loss and gain identification that are often based upon ad hoc algorithms.

In this paper, a new method of aCGH-based instability assessment is presented. The method is based on a pseudolikelihood ratio test (PLRT) and utilizes the aCGH modeling approach presented by Engler et al. (2006). An evaluation of the method is presented through comparison of instability features between meningioma cancer subtypes (Shen et al., 2007) and through simulation studies. Results are compared with those obtained from existing instability metrics. The PLRT method is shown to be more sensitive with regard to distinguishing between different instability types.

1.1 Genetic Instability

Cancer development and progression is a complex process that is influenced by a variety of enviromental and genetic factors. While understanding of the entire etiologic process is not complete, it is generally accepted that genetic instability, introduced through defects in the cytogenic mechanism, leads to the unregulated cell growth that characterizes tumorigenesis (Michor et al., 2004). Such defects may take a variety of forms. For one, genetic alteration can result in the formation of one or more oncogenes. Oncogenes are genes involved in the process of proliferation that when mutated or expressed at abnormally high levels lead to uncontrolled cell division and creation. Alternatively, gene mutation might lead to a change in the actual protein product of a gene. If such a change led to the constant activation of the product, even in the absence of positive signals, an increase in cell division and replication might occur. Genetic alteration might also lead

to the inactivation of tumor suppressor genes. Tumor suppressor genes are genes whose product suppresses proliferation. Inactivation of one or more of these genes might lead to unregulated cell growth. A similar result might be observed if gene mutation leads to an insensitivity to antigrowth signals. The failure of apoptosis, necessary for cell population regulation, is a third possible result of gene mutation.

Such cytogenic defects (and others – see Hanahan and Weinburg, 2000) can be introduced by a variety of genetic events such as whole chromosome loss, segment deletion, segment amplification, translocations, and inversions. There are, in turn, a variety of tools such as FISH, M-FISH, expression analysis, SKY, microsatellite analysis, DNA sequencing, and aCGH that are used to identify occurrences of these varied events. Ideally, results from these methods would be used to ascertain locations of cancer-susceptibility genes along with the specific mutations that lead to unregulated cell growth.

1.2 aCGH-Based Genetic Instability Assessment

Analysis of aCGH-based instability plays an important role in the attempt to understand the events involved in cancer progression. Links between cytogenic mechanism defect and three basic patterns of copy number alteration have been established (see Nowak et al. (2002); Snijders et al. (2003); Raj et al. (2003); Davies et al. (2005); Pinkel and Albertson (2005); Fridlyand et al. (2006)).

First, the total amount of alteration is informative regarding specific cytogenic mechanisms underlying the cancer; cancer progression is often characterized by an increase in the total amount of copy number alteration. Second, the frequency of transition between states such as copy number gain, loss, and no-change (*i.e.*, no gain or loss) may be associated with cancer etiology. For example, whole chromosome loss, or aneuploidy,

is often due to chromsomal instability (CIN). In contrast, tumors with mismatch repair (MMR) deficiencies exhibit few chromosome-wide alterations; alterations in these tumors are typically narrow in scope. Changes in the frequency of transition between states may also reflect differences in cancer stage; a higher frequency of transition is often associated with cancer progression. Two meningioma tumors from the malignant and benign subtypes (Shen et al., 2007), respectively, with different frequencies of transition are shown in Figure 1. Third, the copy number alteration level (*i.e.*, mean log_2 ratio level) may also reflect cancer stage. In Figure 2, two meningioma tumors from the benign and atypical subtypes, respectively, with varying copy number levels are shown. Copy number phenotype is then comprised of these three patterns of alteration. By quantifying and comparing differences in copy number phenotype across cancer subtypes, improvements in the understanding of cancer etiology might be made.

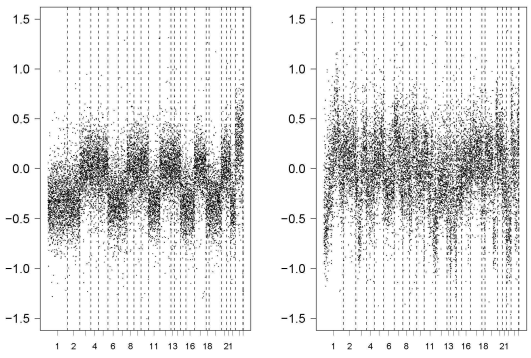

Figure 1: Comparison of malignant (right) and benign (left) menigioma samples exhibiting alterations involving relatively high frequency of transition between states vs. low frequency of transition between states.

1.3 Current Measures of aCGH-Based Instability

Current measures of aCGH-based instability assessment are dependent upon a data preprocessing step in which assignment of clones to either altered (gain, loss) segments

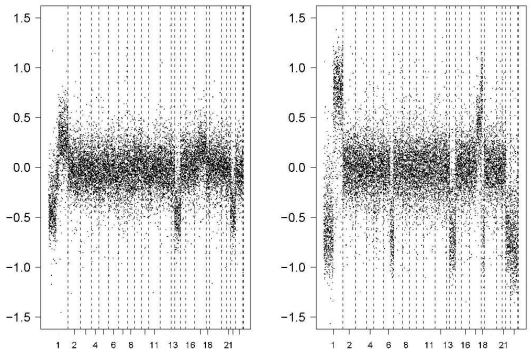

Figure 2: Comparison of benign (right) and atypical (left) menigioma samples exhibiting low-magnitude alterations vs. high-magnitude alterations.

or nonaltered (no change) segments is first conducted. To date, all instability assessment methods have relied upon segmentation-based classification approaches. Such approaches entail the assignment of clones to segments, or sets of contiguous clones, that are classified as gain, loss, or no change. Some authors (*e.g.*, Pollack et al. (2002); Weiss et al. (2003); Aguirre et al. (2004)) have utilized a threshold-based approach for assignment of clones. Clones with absolute log_2 ratio above a given threshold are classified as altered. Contiguous clones in the same threshold region constitute a single segment. Others have employed assignment mechanisms in which data are first segmented into regions of common mean (CBS (Olshen et al., 2004), HMM (Fridlyand et al., 2004), GLAD (Hupe et al., 2004), CLAC (Wang et al., 2005)). Segments are subsequently classified as altered either through use of a threshold or through a method of segment combination (*e.g.*, GLADmerge (Hupe et al., 2004), MergeLevels (Willenbrock and Fridlyand, 2005)).

To date, differences in aCGH-based instability have been assessed through the use of univariate segment-based metrics. In each case, a single number was calculated for each hybridization and comparisions have been made across tumor groups or subtypes.

Primarily, these metrics have been based upon the total number of altered clones (*i.e.*, the

sum of altered segment lengths) per tumor. In some papers, differences (across cancer subtypes) in the average number of alterations per hybridization were merely reported and statistical significance was not assessed (see Martinez et al. (2002); Weiss et al. (2003); Bernardini et al. (2005)). In other papers, comparisons based upon the total number of alterated clones per tumor were made utilizing tests of statistical significance (see Peng et al. (2003); Paris et al. (2004); Jones et al. (2005); Lambros et al. (2005); O'Regan et al. (2006); Wilting et al. (2006); Natrajan et al. (2007)).

Additional univariate segmentation-based metrics have also been employed. Blaveri et al. (2005) and Herzog et al. (2006), for example examined differences in the number of whole chromsome altered segments per tumor across cancer subtypes. Blaveri et al. (2005) also employed both the average number of segments (per tumor) and the average number of chromosomes (per tumor) containing at least two segments as measures of instability.

1.4 Alternative Segmentation-Based Measures of Instability

Notably, none of the previously proposed segmentation-based measures incorporates the magnitude or height of alteration level. However, alteration height differences are an important component of instability assessment; segment height differences are often due to differences in the cytogenic mechanism.

Several potentially useful measures that incorporate segment height are possible. First, the total absolute area under the "curve" might be used where the area created by each segment is its length multiplied by its absolute mean magnitude. The area metric is calculated as the sum of the individual areas. A second possible metric that incorporates segment height makes specific use of the vertical distance between adjacent segments at each breakpoint. The distance metric is calculated as the sum of these distances across

all breakpoints in a given hybridization. Both the area and the distance metrics provide a combined measure of the magnitude and the number of alterations. That is, the measures would provide a higher value for a tumor with a given number of large-magnitude alterations than they would for a tumor with the same number of small-magnitude alterations.

1.5 Limitations of Current Methods

Current methods of genetic instability assessment have several limitations. First, they often fail to distinguish between different types of genetic instability. For example, using the sum of altered segment lengths (*i.e.*, the total number of altered clones) a set of tumors primarily consisting of a few whole-chromosome alterations would not be distinguished from a set consisting of a larger number of small alterations when the total number of average alterations per tumor in both sets is similar. Similarly, no distinction is made using this metric when the difference between sets entails a difference in magnitude or height of altered segments.

Second, results obtained from current metrics are sometimes difficult to interpret. This is due to the fact that while instability is characterized by several distinct features (e.g., magnitude of copy number alteration level, total amount of alteration, frequency of transition between states such as gain and loss), the metrics described in Sections 1.3 and 1.4 are all univariate. Because differences in instability phenotype may be comprised of differences in any combination of these features, interpretation of univariate metrics can be difficult; it is often not clear which combination of feature differences results in an identified difference (or lack of identified difference) between cancer subtypes. Hence, use of a single univariate is limited in this setting.

A third limitation of some current methods is that they require a second stage of analysis following ad hoc methods of DNA sequence copy number alteration identification. For example, most methods of overall instability assessment have identified altered segments through the use of thresholds (Martinez et al. (2002); Weiss et al. (2003); Peng et al. (2003); Jones et al. (2005); O'Regan et al. (2006)). Alternatively, Fridlyand et al. (2006) have used the MergeLevels segment combination procedure as a foundation for genetic instability assessment. The MergeLevels approach does not require ad hoc selection of a threshold. However, the MergeLevels approach may be limited in its ability to detect small regions of alteration (see Engler et al., 2006) and may hence lead to insensitive measures of instability.

2 Methods: A Model-Based Method of Instability Assessment

2.1 aCGH Classification

A few authors have proposed model-based approaches for aCGH data preprocessing (*e.g.*, Fridlyand et al. (2004); Engler et al. (2006); Picard et al. (2007)). The parameters in the Fridlyand et al. and Picard et al. models, however, do not correspond to natural instability metrics and have not been used for instability assessment. Engler et al. (2006) proposed a likelihood-based aCGH classfication method that offers a formal modeling framework for instability assessment. The approach is based on a three-state Gaussian mixture model (for loss, no-change, and gain events), with a hidden Markov dependence structure, and with random effects to allow for both inter- and intra-tumoral clonal variation. For computational ease, estimation of model parameters and of posterior event probabilities is based on a pseudolikelihood function (Arnold and Strauss (1991)) in which dependen-

cies within each triplet of data are modeled and the likelihood contribution of each triplet are multiplied together. The method provides posterior probabilities of gain, loss, and no change, supplying quantitative evidence of alteration events. Through data analysis and simulation studies, Engler et al. (2006) showed that the method more accurately classifies small regions of alterations than segmentation-based methods. The authors also found that the method has greater accuracy when intratumoral clonal variation is present.

Several elements of this modeling approach are particularly well-suited to instability assessment. For one, an important component of instability phenotype is the magnitude or height of alteration levels. Through estimation of the underlying state mean parameters in the mixture model (*i.e.*, means for gain, no change, and loss), the approach provides such information. Second, both the overall amount of alteration and the frequency of transition between states of loss, no-change, and gain are also important components of the instability pattern. By incorporating the Markov dependence structure, the model quantifies these features through the marginal state and transition probabilities.

2.2 Pseudolikelihood Ratio Test (PLRT)

As described in Section 1.5, univariate measures provide an incomplete summary of differences in instability phenotype. A test that globally assesses differences in copy number segment length, number, and height is of interest. The approach proposed by Engler et al. (2006) affords formal hypothesis testing to make such assessment between tumor sets through evaluation of differences in state means, marginal probabilities, and transition probabilities. Under the null hypothesis, different cancer subtypes are represented by a single set of mean parameters, marginal probabilities, and transition probabilities. A variety of alternative hypotheses are possible. For example, if it were of interest to test for differences in the magnitude of transition events (*i.e.*, heights of altered segments) be-

tween two groups, the mean parameters would be allowed to vary between two groups. Alternatively, the marginal and transition probabilities would be allow to vary between two groups if it were of interest to test whether the total amount of alteration and frequency of transition between states differed between the two groups. Under the global alternative hypothesis of any difference between groups, the two mean parameters along with the marginal and transition probabilities are allowed to vary between groups.

Let $(\mathbf{y}_1, \ldots, \mathbf{y}_N)$ represent vectors of log_2 ratios from all N hybridizations. Let S be the set of all possible unique tripletwise indicator vectors s for a hybridization with J clones; each $s \in S$ is of length J and consists of three consecutive 1's, and 0's for the remaining $J-3$ entries. Finally, let $\mathbf{y}_i^{(s)}$ be the subvector of $\mathbf{y}_i$ that corresponds to the elements of s equal to 1. The log pseudolikelihood proposed by Engler et al. (2006) is

$$pl = \sum_{i=1}^{N} \sum_{s \in S} \ln f_s(\mathbf{y}_i^{(s)}; \boldsymbol{\Theta}, \mathbf{C}_s), \tag{2.1}$$

where f_s is the likelihood contribution for each triplet indicated by s, $\mathbf{C}_s$ is the vector consisting of the true underlying states for the triplet, and $\boldsymbol{\Theta}$ is the vector of model parameters including the state mean parameters and the marginal and transition probabilities.

Let the complete parameter vector be represented as $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{\delta})$, where $\boldsymbol{\delta} = \boldsymbol{\delta}_0$ is the subvector of interest and where $\boldsymbol{\alpha}$ is a nuisance parameter vector. For example, one test of interest entails examining whether the means of DNA copy number gain and loss for one subgroup differ from the means of gain and loss of a second subgroup. Let $\boldsymbol{\delta} = (\delta_G, \delta_L)$ and let $\boldsymbol{\alpha} = (\mu_G, \mu_L, \mathbf{v})$, where $\mathbf{v}$ represents the vector containing the remaining model parameters. Under H_0, μ_G and μ_L represent the mean of DNA copy number gain and loss,respectively, for the entire set of hybridizations. Under the alternative hypothesis, H_A, μ_E is the mean of event E, $E = G, L$, in group 1 and $(\mu_E + \delta_E)$ is the mean of event E in group 2. Under H_0, $\boldsymbol{\delta} = \mathbf{0}$.

A formal comparison of H_0 and H_A can then be conducted using the pseudolikelihood ratio test statistic.

$$G^2 = 2[pl(\tilde{\boldsymbol{\Theta}}) - pl(\tilde{\boldsymbol{\alpha}}(\boldsymbol{\delta}_0), \boldsymbol{\delta}_0)], \tag{2.2}$$

where the function $pl(.)$ is defined in (2.1), $\tilde{\boldsymbol{\Theta}}$ is the vector of maximum pseudolikelihood estimates, and where $\tilde{\boldsymbol{\alpha}}(\boldsymbol{\delta}_0)$ is the vector of maximum pseudolikelihood estimates for $\boldsymbol{\alpha}$ under H_0. The distribution of (2.2) is asymptotically equivalent to that of a weighted sum of χ^2 random variables (Geys et al., 1999). Following calculation of the weights, testing can be conducted by comparing the observed value of (2.2) to a large number of draws (*e.g.*, $100,000$) from its null distribution. Performance of the pseudolikelihood ratio test (PLRT) is evaluated in Sections 3.1 and 3.2.

Use of alternative formal testing procedures is also possible. Geys et al. (1999), for example, proposed use of pseudolikelihood-based score tests and Wald tests. However, through simulation studies, the authors found that the pseudoscore test and Wald test had lower power than their likelihood-based counterparts and suggested use of the pseudolikelihood ratio test. Formal subgroup comparison is also possible through use of a permutation test in which the group labels are randomly permuted among hybridizations. For each set of label assignments, the pseudolikelihood ratio test statistic, G^2, is computed. The value of G^2 under the true hybridization assignment is then compared to its permutation distribution and a p-value is calculated. Of note, the permutation test can be useful in examining the effect of model assumptions. Assessment of PLRT model assumptions using a small permutation test is described in Section 3.1. However, because aCGH data sets are typically quite large, this approach is not practical because of its computational burden.

2.3 Software

Analyses were performed using the R software package (http://www.r-project.org). The R implementation of the pseudolikelihood ratio tests presented in this paper is available at http://www.biostat.harvard.edu/~betensky/
papers.html.

3 Results

3.1 Meningioma Data

Meningiomas are a common type of intracranial tumor. Shen et al. (2007) analyzed genomic DNA from a cohort of sporadic solitary menigiomas using cDNA microarray chips. The 72 samples were previously classified as benign ($n = 34$), atypical ($n = 25$), or malignant ($n = 13$). It is of interest to test for differences in genetic instability between pathological grades.

Of note, the meningioma data contained considerable noise; the estimated variance of the no-change events was twice that of the glioma data set analyzed by Engler et al. (2006). In part, the noise was attributed to single-clone outliers. The investigators of the meningioma study determined that these outliers were due to experimental variability rather than underlying biologic or etiologic factors. This determination was based on examination of several samples generated on both the cDNA microarray platform and the higher resolution long oligo(60mer) based microarray. Hence, all analyses were conducted on smoothed data in which outliers were removed. Outliers were identified as those clones whose absolute difference in magnitude from both of its two adjacent neighboring clones

exceeded 2σ. The estimate of the no-change variance parameter was used for σ.

The noise of the data, even after smoothing, also led to difficulties in estimation of the random effect variance components of the Engler et al. (2006) model. In simulation studies (unpublished), it was observed that when substantial noise is present, this model has difficulty distinguishing between random effect variance components and the variance of the log_2 ratios. Hence, for the following analyses, the random effects portion of the model was not utilized.

Several segmentation-based instability metrics were calculated to assess differences in instability type between the three cancer subtype groups. First, for each hybridization, the sum of altered segment lengths (SSL) was calculated using a specified threshold t. Clones with an absolute magnitude greater than threshold t were categorized as alterations (*i.e.*, gains or losses). The threshold value $t = 0.25$ was identified by Shen et al. (2007) as one appropriate for this particular data set. To check sensitivity to threshold choice in instability assessment, values of 0.20 and 0.30 were also used in separate analyses. An additional analysis was conducted using the MMAD method (see Paris et al., 2004; Rossi et al., 2005) in which a nonparametric estimate, $\tilde{\sigma}$, of the standard deviation for each hybridization is obtained through use of the segmentation results. For each segment, the median absolute deviation (MAD) of the log_2 ratios (belonging to that segment) from the segment mean is calculated. The MMAD estimate of σ is then calculated as the median MAD across all segments for a given hybridization. For this analysis, all clones belonging to segments with means above $t = 2\tilde{\sigma}$ or below $t = -2\tilde{\sigma}$ were categorized as gains or losses, respectively. Across tumor groups, the median (across hybridizations in each group) MMAD estimates for $2\tilde{\sigma}$ are $2\tilde{\sigma}_{\text{benign}} = 0.232$, $2\tilde{\sigma}_{\text{atypical}} = 0.241$, and $2\tilde{\sigma}_{\text{malignant}} = 0.244$.

Second, the CBS approach (Olshen et al., 2004) was used as the basis for the two segmentation metrics described in Section 1.4. Both the "area" metric and the "distance"

Table 1: *Results (*p*-values) from Wilcoxon rank-sum tests. Pairwise comparisons are made between benign (B, n=34), atypical (A, n=25), and malignant (M, n=13) groups.*

Metric	B-A	A-M	B-M
SSL[1] (t=0.20)	0.041	0.104	0.001
SSL (t=0.25)	0.031	0.091	0.001
SSL (t=0.30)	0.0163	0.091	0.001
SSL (MMAD)	0.001	0.564	0.001
area	0.001	0.125	< 0.001
distance	0.348	0.023	0.007

1 SSL: sum of segment lengths (*i.e.*, total number of altered segments)

metric were calculated for each hybridization using the combined CBS segments of the MergeLevels (Willenbrock and Fridlyand, 2005) approach.

For each of the above metrics, a Wilcoxon rank-sum test was used to test for significant differences between groups. Table 1 contains the Wilcoxon rank-sum test results for each of the two-group comparisons using each of the instability metrics. The group comparisons are benign vs. atypical (B-A), atypical vs. malignant (A-M), and benign vs. malignant (B-M).

All methods of segmentation-based instability assessment identify a signficant difference (at the 0.05 level) between the two most clinically distinct subtypes of benign and malignant. Additionally, with the exception of the distance metric, all methods also identify a significant difference in amount of alteration between the benign and atypical subtypes. In the comparison between the atypical and malignant subtypes, only the distance metric detects a difference between the two groups, though as a mathematical combination of two features it is difficult to interpret. Despite the general agreement across metrics, it is difficult to assess the nature of the instability phenotype differences between the two groups in terms of biologically meaningful parameters. That is, it is unclear whether the differences are due to variations in the total number of alterations, in the frequency of

transition between states, or in the magnitude of alteration events.

The distance metric is based on the mean levels of segments and the number of segments, which itself is a function of both the frequency of alteration between states and the overall amount of alteration. Thus, results obtained from the distance metric are difficult to interpret. However, the discrepancy between five of the metrics (four threshold metrics and the area metric) and the distance metric for the benign-atypical and atypical-malignant comparisons suggests additional complexity in instability differences that is not easily explained by univariate measures.

Analysis of the meningioma data was also conducted using tests using the pseudolikelihood approach of Engler et al. (2006). Parameter estimates for the three subtype groups are listed in Table 2. The symbols L, 0, and G represent copy number loss, no-change, and copy number gain, respectively. The mean for event A is represented by μ_A. The notation $P(A)$ represents the marginal probability of event A and $P(A|B)$ represents the conditional probability of event A given event B (*i.e.*, the transition probability). Several differences between groups are apparent. There appears to be a downward shift in μ_L from benign to atypical to malignant. This may be reflective of advances in cancer stage. Additionally, there is a upward trend in the amount of copy number gain, represented by the marginal probability of gain $P(G)$. Notable changes across tumor types are a decrease in $P(L|G)$ and an increase in $P(G|G)$. These are related to the corresponding increases in $P(G)$.

Formal comparisons between groups were made using the PLRTs. For each two-group comparison, four PLRTs were conducted. Results are displayed in Table 3. First, differences in magnitude of alteration levels were assessed using means-tests, which test the null hypothesis that the mean levels of gain and loss are the same across both groups. Second, differences in frequencies of changes between states were assessed using

Table 2: *Parameter estimation results from the pseudolikelihood approach for subtype groups benign (n=34), atypical (n=25), and malignant (n=13).*

Metric	Benign (B)	Atypical (A)	Malignant (M)
μ_L	-0.349	-0.524	-0.537
μ_0	0.008	-0.012	-0.013
μ_G	0.461	0.431	0.448
σ^2	0.029	0.032	0.039
$P(L)$	0.097	0.100	0.108
$P(0)$	0.859	0.845	0.765
$P(G)$	0.044	0.055	0.127
$P(L\|L)$	0.838	0.912	0.899
$P(0\|L)$	0.150	0.084	0.095
$P(G\|L)$	0.012	0.004	0.006
$P(L\|0)$	0.017	0.001	0.013
$P(0\|0)$	0.973	0.982	0.981
$P(G\|0)$	0.010	0.008	0.006
$P(L\|G)$	0.027	0.008	0.005
$P(0\|G)$	0.195	0.123	0.036
$P(G\|G)$	0.778	0.869	0.959

transition-probability-tests, which test the null hypothesis that the transition probabilities are the same across both groups. Third, differences in overall amount of gain and loss events were assessed using state-probability-tests, which test the null hypothesis that the marginal probabilities of alteration do not vary across groups. Global-tests were also conducted, which consisted of joint comparisons of means, transition probabilities, and state probabilities.

Table 3: *Results (*p*-values) from pseudolikelihood ratio tests (PLRT). Comparisons are made between benign (B, n=34), atypical (A, n=25), and malignant (M, n=13) groups.*

Metric	B-A	A-M	B-M
means-test	0.0005	0.5369	0.0092
transition-probability-test	0.0732	0.3051	0.2159
state-probability-test	0.0377	0.1344	0.0133
global-test	0.0313	0.2550	0.0169

The benign and malignant groups differ both in the magnitude of alteration levels (p=0.0092) and in the overall amount of gain and loss (p=0.0133). The two groups do not differ with regard to frequency of transition between event states (p=0.2159). The two groups were also found to differ significantly under the global test, in which state means, transition probabilities, and marginal state probabilities were allowed to vary across groups. Hence, the benign and malignant groups differ primarily in magnitude of alteration levels and in the overall amount of gain and loss. Notably, these results are more informative than those provided by the segmentation-based metrics in which assessment of specific instability phenotype differences is not feasible.

The benign and atypical groups differ both in the magnitude of alteration levels (p=0.0005) and in the overall amount of gain and loss (p=0.0377). The two groups are marginally significantly different in frequency of transition between event states (p=0.0732). They were also found to differ significantly under the global test (p=0.0313). Thus, as in the benign-malignant comparison, the benign and atypical groups differ primarily in magnitude of alteration levels and in the overall amount of gain and loss. Again, the difference in specificity between the PLRT results and the segmentation-based results is apparent.

No significant differences were detected between the atypical and malignant groups. It should be noted, however, that the atypical-malignant test results are based on a comparison entailing the two smallest sample sizes ($n = 25$ and $n = 13$). There is, in fact, some evidence that there may be true differences. First, results obtained from the distance metric (see Table 1) suggest some type of difference between groups. Second, the group parameter estimates in Table 2 suggest that there is a difference between the atypical and malignant groups in probabilities of no change and gain. It is likely that with larger sample sizes, differences between the atypical and malignant groups would be observed.

3.2 Simulations

Simulation studies were conducted to compare the performance various instability metrics under several data scenarios. For each scenario, simulated data was generated using a Markov dependence structure in which overall state means (loss, no change, gain), overall variance, marginal probabilities, and transitional probabilities were specified. Simulated log_2 ratios were distributed normally about the state means. For the first simulation study, transition probabilities varied across groups and state means and marginal probabilities did not. For this simulation, 200 data sets were generated, consisting of two groups of $n = 8$ hybridizations per group with 500 data points per hybridization. Results of this simulation study are presented in Tables 4 and 5 under the heading "Scenario 1". For the second simulation study, state means and the transition probabilities varied between the two groups and the overall amount of gain and loss did not. Because this scenario entailed a more precise set of circumstances (*i.e.*, no difference in the average sum of vertical distances between segments over both groups) and fewer constraints (*i.e.*, both transition probabilities and mean parameters were allowed to vary), an increased sample size was utilized: 200 data sets were generated, consisting of two groups of $n = 15$ hybridizations per group with 1000 data points each. Results of this second scenario are provided in Tables 4 and 5 under the heading "Scenario 2".

In Scenario 1, the overall amount of alteration (*i.e.*, the total number of clones categorized as alterations) and the mean levels of alteration do not vary between groups. The area metric and the threshold-based sum of segment lengths (SSL) will, on average, fail to distinguish between groups when the sole difference is in the frequency of transition between event states. Alternatively, the distance metric will be able to distinguish between groups in this scenario since an increase in the frequency of transition between states (holding all else constant) will result in more segments and a greater sum distance between segments. The power estimates for these metrics at $\alpha = 0.05$ are listed in Table

4 and conform to these expectations. The low power (0.065) of the SSL metric is due to the fact that the metric is threshold based and does not detect differences between groups unless the overall amount of transition varies.

The Type-I error for the PLRT means-test is 0.050 (Table 5); under Scenario 1 no true difference exists between the state means of the two groups. For the transition-probability-test, high power (0.855) is observed in this data scenario. Since the only difference between the two groups is in the frequency of transition between events, this is to be expected. The global-test has a lower power (0.325) than the transition-probability-test in this scenario. This is due to the fact that the state means and marginal probabilities do not vary between the two groups. As in the data analysis, the PLRT test results provide more information as to the specific type of instability differences between the two groups than do the univariate metric based tests.

For Scenario 2, the area metric is more powerful (0.270) than the distance metric (0.045); for this scenario, the differences in mean level and transition probabilities are such that, on average, the area metric identifies a difference between groups but the distance metric does not. The SSL test again has low power (0.085) in this case due to the fact that the overall amount of alteration does not vary between the two groups. The PLRT means-test and global-test have the highest power (1.000) of the three tests. The lower power of the transition-probability-test is seemingly due to model misspecification; under the alternative hypothesis for this test the state means are constrained to be equal while the transition probabilities are allowed to vary. Again, results of the PLRTs are easily interpretable and provide a greater amount of information with regard to instability type differences between subgroups.

Table 4: *Estimate of power ($\alpha = 0.05$) of segmentation-based tests under two simulated data scenarios*

Metric	Scenario 1	Scenario 2
SSL (MMAD)	0.0650	0.0850
area	0.0650	0.2700
distance	0.8400	0.0450

Table 5: *Estimate of power ($\alpha = 0.05$) of pseudolikelihood ratio tests (PLRT) under two simulated data scenarios*

Metric	Scenario 1	Scenario 2
means-test	0.0500	1.000
transition-probability-test	0.8550	0.4500
global-test	0.3250	1.000

4 Discussion

To date, all methods of instability assessment are based on univariate measures. However, instability phenotypes are multidimensional and are comprised of several features such as magnitude of copy number alteration level, total amount of alteration, and frequency of transition between states such as gain and loss. Differences in instability phenotype may be due to differences in any combination of these features. Because of this potential complexity, interpretation of univariate metrics is difficult; it is often not clear which combination of feature differences results in an identified difference (or lack of identified difference) between cancer subtypes. Hence, use of a single univariate is limited in this setting.

The PLRT, alternatively, provides clear assessment of instability differences. Through use of the Engler et al. (2006) approach, differences in magnitude of copy number alteration level, total amount of alteration, and frequency of transition between states can be assessed directly through tests involving mean parameters, marginal state probabilities,

and transition probabilities, respectively. Tests for differences in specific combinations of these instability phenotype features can likewise be conducted.

Conceivably, simple segmentation-based metrics might be formulated to mimic the selection of tests available through the PLRT. For example, a test of means might be conducted by comparing the average mean segment magnitudes between cancer subtypes. A test might also be construed to compare frequency of transition between states by assessing differences in segment lengths between subtypes. However, such results of such tests would be sensitive to threshold selection. Also, they involve a second stage of analysis following estimation and signal processing and are not part of a unified framework for analysis. Alternatively, the PLRT provides a model-based tool for estimation and testing of aCGH-based instability phenotype differences, with easily interpretable results.

References

[1] Aguirre,A., Brennan,C., Bailey,G., Sinha,R., Feng,B., Leo,C., Zhang,Y., Zhang,J., Gans,J., Bardeesy,N., Cauwels,C., Cordon-Cardo,C., Redston,M., DePinho,R., Chin,L. (2004) High-resolution characterization of the pancreatic adenocarcinoma genome. *PNAS* **24**, 9067-9072.

[2] Arnold,B.C., Strauss,D. (1991) Pseudolikelihood estimation: some examples. *Sankhya* Ser. B, **53**, 233-243.

[3] Bernardini,M., Lee,C.H., Beheshti,B., Prasas,M., Albert,M., Marrano,P., Begley,H., Shaw,P., Covens,A., Murphy,J., Rosen,B., Minkin,S., Squire,J.A., Macgregor,P.F. (2005) High-resolution mapping of genomic imbalance and identification of gene expression profiles association with differential chemotherapy response in serous epithelial ovarian cancer. *Neoplasia* **7**, 603-613.

[4] Blaveri,E., Brewer,J.L., Roydasgupta,R., Fridlyand,J., DeVries,S., Koppie,T., Pejavar,S., Mehta,K., Carroll,P. (2005) Bladder cancer stage and outcome by array-based comparative genomic hybridization. *Clinical Cancer Research* **11**, 7012-7022.

[5] 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.K. (2006) Array comparative genomic hybridization reveals genomic copy number changes associated with outcome in diffuse large B-cell lymphomas. *Blood* **107**, 2477-2485.

[6] Davies,J.J., Wilson,I.M., Lam,W.L. (2005) Array CGH technologies and their applications to cancer genomes. *Chromosome Research* **13**, 237-248.

[7] Engler,D.A., Mohapatra,G., Louis,D.N., Betensky,R.A. (2006) A pseudolikelihood approach for simultaneous analysis of array comparative genomic hybridizations. *Biostatistics* **7**, 399-421.

[8] Fridlyand,J., Snijders,A., Pinkel,D., Albertson,D., and Jain,A. (2004) Hidden Markov models approach to the analysis of array CGH data. *Journal of Multivariate Analysis* **90**, 132-153.

[9] Fridlyand,J., Snijders,A., Ylstra,B., Li,H., Olshen,A., Segraves,R., Dairkee,S., Tokuyasu,T., Ljung,B.M., Jain,A.N., McLennan,J., Ziegler,J., Chin,K., Devries,S., Feiler,H., Gray,J.W., Waldman,F., Pinkel,D., Albertson,D.G. (2006) Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* **6**:96.

[10] Geys,H., Molenberghs,G., Ryan,L.M. (1999) Pseudolikelihood modeling of multivariate outcomes in developmental toxicology. *Journal of the American Statistical Association* **94**, 734-745.

[11] Hanahan,D., and Weinberg,R.A. (2000) The hallmarks of cancer. *Cell* **100**, 57-70.

[12] Herzog,C.R., Desai,D., Amin,S. (2006) Array CGH analysis reveals chromosomal aberrations in mouse lung adenocarcinomas induced by the human lung carcinogen

4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone. *Biochemical and Biophysical Research Communications* **341**, 856-863.

[13] Hodgson,G., Hager,J.H., Volik,S., Hariono,S., Wernick,M., Moore,D., Albertson,D.G., Pinkel,D., Collins,C., Hanahan,D., and Gray,J.W. (2001) Genome scanning with array CGH deliniates regional alternatives in mouse islet carcinomas. *Nature Genetics* **29**, 459-464.

[14] Hupe,P., Stransky,N., Thiery,J.P., Radvanyi,F., Barillot,E. (2004) Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics* **20**, 3413-3422.

[15] Jones,A.M., Douglas,E.J., Halford,S.E.R., Fiegler,H., Gorman,P.A., Roylance,R.R., Carter,N.P., Tomlinson,I.P.M. (2005) Array-CGH analysis of microsatellite-stable, near-diploid bowel cancers and comparison with other types of colorectal carcinoma. *Oncogene* **24**, 118-129.

[16] Lambros,M.B.K., Fiegler,H., Jones,A., Gorman,P., Roylance,R.R., Carter,N.P., Tomlinson,I.P.M. (2005) Analysis of ovarian cancer cell lines using array-based comparative genomic hybridization. *Journal of Pathology* **205**, 29-40.

[17] Martinez-Climent,J.A., Alizadeh, A.A., Segraves,R., Bleasa,D., Rubio-Moscardo,F., Albertson,D.G., Garcia-Conde,J., Dyer, M.J.S., Levy,R., Pinkel,D., Lossos,I.S. (2002) Transformation of follicular lymphoma to diffuse large cell lymphoma is associated with a heterogeneous set of DNA copy number and gene expression alterations. *Blood* **101**, 3109-3117.

[18] Michor,F., Iwasa,Y., Nowak,M. (2004) Dynamics of cancer progression. *Nature Reviews Cancer* **4**, 197-205.

[19] Natrajan,R., Little,S.E., Sodha,N, Reis-Filho,J.S, Mackay,A., Fenwick,K., Ashworth,A., Perlman,E.J., Dome,J.S., Grundy,P.E., Pritchard-Jones,K., Jones,C. (2007)

Analysis by array CGH of genomic changes associated with the progression of relapse of Wilms' tumour. *Journal of Pathology* **2007**, 52-59.

[20] Nowak,M.A. Komarova,N.L., Sengupta,A., Jallepalli,P.V., Shih,L., Vogelstein,B., Lengauer,C. (2002) The role of chromosomal instability in tumor initiation. *PNAS* **99**, 16226-16231.

[21] O'Regan,E.M., Toner,M.E., Smyth,P.C., Finn,S.P., Timon,C., Cahill,S., Flavin,R., O'Leary,J.J., Shelis,O. (2006) Distinct array comparative genomic hybridization profiles in oral squamous cell carcinoma occuring in young patients. *Head & Neck* **28**, 330-338.

[22] Olshen,A.B., Venkatraman,E.S., Lucito,R., Wigler,M. (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **4**, 557-572.

[23] Paris,P.L., Andaya,A., Fridlyand,J., Jain,A.N., Weinberg,V., Kowbel,D., Brebner,J.H., Simko,J., Watson,J.E.V., Volik,S., Albertson,D.G., Pinkel,D., Alers,J.C., van der Kwast,T.H., Vissers,K.J., Schroder,F.H., Wildhagen,M.F., Febbo,P.G., Chinnaiyan,A.M., Pienta,K.J., Carroll,P.R., Rubin,M.A., Collins,C., van Dekken,H. (2004) Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Human Molecular Genetics* **13**, 1303-1313.

[24] Peng,D., Sugihara,H., Mukaisho,K., Tsubosa,Y., Hattori,T. (2003) Alterations of chromosomal copy number during progression of diffuse-type gastric carcinomas: metaphase- and array-based comparative genomic hybridization analysis of multiple samples from individual tumours. *Journal of Pagimpthology* **201**, 439-450.

[25] Picard,F., Robin,S., Lebarbier,E., Daudin,J. (2007) A segmentation/clustering method for the analysis of array CGH data. *Biometrics* In press.

[26] Pinkel,D. and Albertson,D.G. (2005) Comparative genomic hybridization. *Annu Rev Genomics Hum Genet* **6**, 331-354.

[27] Pollack,J.R., Sorlie,T., Perou,C.M., Rees,C.A., Jeffrey,S.S., Lonning,P.E., Tibshirani,R., Botstein,D., Borresen-Dale,A., and Brown,P.O. (2002) Microarray analysis reveals a major direct role of DNA copy number alternation in the transcriptional program of human breast tumors *PNAS* **99**, 12963-12968.

[28] Rajagopalan,H., Nowak,M.A., Vogelstein,B., Lengauer,C. (2003) The significance of unstable chromosomes in colorectal cancer. *Nature Reviews Cancer* **3**, 695-701.

[29] Rossi,M.R., Gaile,D., LaDuca,J., Matsui,S.I., Conroy,J., McQuaid,D., Chervinsky,D., Eddy,R., Chen,H.S., Barnett,G.H., Nowak,N.J., Cowell,J.K. (2005) Identification of consistent novel submegabase deletions in low-grade oligodendrogliomas using array-based comparative genomic hybridization. *Genes, Chromosomes & Cancer* **44**, 85-96.

[30] Shen,Y., Nunes,F., Stemmer-Rachamimov,A., James,M., Mohapatra,G., Plotkin,S., Betensky,R.A., Engler,D.A., Roy,J., Ramesh,V., Gusella,J. (2007) Genomic profiling of meningiomas by array comparative genomic hybridization. *Cancer Research*, In press.

[31] Snijders,A.M., Fridlyand,J., Mans,D.A., Segraves,R., Jain,A.N., Pinkel,D., Albertson,D.G. (2003) Shaping of tumor and drug-resistant genomes by instability and selection. *Oncogene* **22**, 4370-4379.

[32] Tagawa,H., Karnan,S., Suzuki,R., Matsuo,K., Zhang,X., Ota,A., Morishima,Y., Nakamura,S., Seto,M. (2005) Genome-wide array-based CGH for mantle cell lymphoma: identification of homozygous deletions of the proapoptotic gene BIM. *Oncogene* **24**, 1348-1358.

[33] Varma,G., Varma,R., Huang,H., Pryshchepava,A., Groth,J., Fleming,D., Nowak,N.J., McQuaid,D., Conroy,J., Mahoney,M., Moysich,K., Falkner,K.L., Geradts,J. (2005) Array comparative genomic hybridization (aCGH) analysis of premenopausal breast cancers from a nuclear fallout area and matched cases from Western New York. *British Journal of Cancer* **93**, 699-708.

[34] Wang,P., Young,K., Pollack,J., Narasimhan,B., and Tibshirani,R. (2005) A method for calling gains and losses in array CGH data. *Biostatistics* **6**, 45-58.

[35] Weiss,M.M., Kuipers,E.J., Postma,C., Snijders,A.M., Stolte,M., Vieth,M., Pinkel,D., Meuwissen,S.G.M., Albertson,D., Meijer,G.A. (2003) Genome wide array comparative genomic hybridization analysis of premalignant lesions of the stomach. *Mol. Pathol.* **56**, 293-298.

[36] Willenbrock,H. and Fridlyand,J. (2005) A comparison study: applying segmentation to array CGH data for downstream analyses. *Bioinformatics* **21** 4084-4091.

[37] Wilting,S.M., Snijders,P.J.F., Meijer,G.A., Yistra,B., van den IJssel,P.R.L.A., Snijders,A.M., Albertson,D.G., Coffa,J., Schouten,J.P., van de Wiel,M.A., Meijer,C.J.L.M., Steenbergen,R.D.M. (2006) Increased gene copy numbers at chromosome 20q are frequent in both squamous cell carcinomas and adenocarcinomas of the cervix. *Journal of Pathology* **209**, 220-230.