2

Methodology of Microarray Data Analysis

Xiaofeng Zhou, Hiroshi Egusa, Steven W. Cole, Ichiro Nishimura, and David T.W. Wong

Introduction

Microarrays have revolutionized biological research by allowing researchers to study the expression of thousands of genes simultaneously for the first time (DeRisi *et al.,* 1996; Schena *et al.,* 1995). Although biomedical investigators have quickly adopted this powerful new research tool, accurate analysis and interpretation of the data have presented an immense challenge in the fields of statistical science and data mining. To make matters worse, microarray technology has rapidly expanded from simply a ribonucleic acid (RNA) expression profiling method to many different applications, including genome-wide single nucleotide polymorphism (SNP) genotyping (Lindblad-Toh *et al.,* 2000a; Lindblad-Toh *et al.,* 2000b; Mei *et al.,* 2000), comparative genomic hybridization (CGH) (Albertson and Pinkel, 2003; Pollack *et al.,* 1999), and various types of protein arrays (Cahill and Nordhoff, 2003). A variety of data analysis tools have been developed to accommodate the various applications for microarray analysis. In this article, we survey some common analytical strategies for expression analysis, which can be potentially adapted to most microarray applications.

The major steps involved in microarray data analysis are as follows: 1) microarray image acquisition and raw data generation, 2) data normalization and transformation, 3) classification and exploratory data analysis, and 4) post-analysis follow-up and validation. The first step, microarray image acquisition and raw data generation, is heavily platform dependent. Regardless of the approach chosen, the arrays are scanned after hybridization. Independent grayscale images, typically 16-bit tiff (tagged information file format) files, are generated for each sample to be analyzed. Image analysis software is then used to identify arrayed spots and measure the relative fluorescence intensities for each element. There are many commercial and freely available software packages for image quantitation. (Go to microarray.genetics.ucla.edu/public/softwarelinks for a list of microarray analysis software.) Although there are differences between various imaging softwares, most give high-quality, reproducible measures of hybridization intensities.

For the purpose of our discussion, we will ignore the particular microarray platform used, the type of measurement reported (mean, median, or integrated intensity for various complementary deoxyribonucleic

Handbook of Immunohistochemistry and *in situ* Hybridization of Human Carcinomas, Volume 3: Molecular Genetics, Liver Carcinoma, and Pancreatic Carcinoma

acid [cDNA] microarray, or the average difference for Affymetrix GeneChips), the background correction performed, or spot-quality assessment and trimming used. As our starting point, we will assume that for each biological sample we assay, we have a high-quality measurement of the intensity of hybridization for each gene element on the array.

Microarray Data Normalization and Transformation

The hypothesis underlying microarray analysis is that the measured intensity for each gene element on the array represents its relative expression level. Biologically relevant expression patterns usually are compared between different states on a gene-by-gene basis. However, before the expression levels can be compared appropriately, a number of transformations must be carried out on the data to eliminate questionable or low-quality measurements and to adjust the measured intensities to facilitate comparisons (Quackenbush, 2002).

Expression Ratios

Most microarray experiments investigate the relative differences in gene expression, and a straightforward approach to this problem involves screening genes based on the ratio of their expression intensity across samples. Although ratios provide an intuitive measure of expression changes, they have the disadvantage of treating up- and down-regulated genes differently. Genes that up-regulated by a factor of 2 have an expression ratio of 2, whereas those down-regulated by the same factor have an expression ratio of (–0.5). The most widely used alternative transformation of the ratio is the logarithm base 2, which has the advantage of producing a continuous spectrum of values and treating up- and down-regulated genes in a similar fashion: $\log_2(1) = 0$, $\log_2(2) = 1$, $\log_2(1/2) = -1$, $\log_2(4) = 2$, $\log_2(1/4) = -2$, and so on. This log transformation treats the expression ratios (or their reciprocals) symmetrically, so that a gene up-regulated by a factor of 2 has a $\log_2$(ratio) of 1, a gene down-regulated by a factor of 2 has a $1\log_2$(ratio) of −1, and a gene expressed at a constant level (with a ratio of 1) has a $\log_2$(ratio) equal to zero. Most statistical analyses of differential expression rely on log-transformed intensity values. Log transformation should only be carried out after noise-suppression measures are taken (because measurement error is often linear in the original assay metric but not in the logfold change metric).

Normalization

A critical step for microarray data analysis is normalization, which adjusts the individual hybridization intensities to balance them appropriately so that meaningful biological comparisons can be made. There are a number of reasons why data must be normalized, including unequal quantities of starting RNA, differences in labeling, hybridization, or detection efficiencies between the fluorescent dyes used, and systematic biases in signal measurement accuracy and sensitivity. Conceptually, normalization is similar to adjusting expression levels measured by northern analysis or quantitative real-time polymerase chain reaction (PCR) relative to the expression of one or more housekeeping genes whose levels are assumed to be constant between samples. Many approaches have been used for microarray normalization, including total intensity-based normalization, Lowess normalization, model-based normalization, linear regression analysis (Chatterjee and Price, 1991), log centering, rank invariant methods (Tseng *et al.,* 2001), and ratio statistics (Chen *et al.,* 1997; Chen *et al.,* 2002). We will describe the three most commonly used methods here: total intensity-based normalization, Lowess normalization, and model-based normalization.

Total Intensity-Based Normalization

Total intensity-based normalization is the most common normalization approach, and it is based on several simple assumptions. The first is that we are starting with equal quantities of RNA for the two (or more) samples to be compared. Given that there are millions of individual RNA molecules in each sample, we will assume that the average mass of each molecule is approximately the same, and that, consequently, the number of molecules in each sample is also the same. The second assumption is that the arrayed elements represent a random sampling of the genes in the genome. This is important because we will also assume that the arrayed elements randomly interrogate the two RNA samples. If the arrayed genes are selected to represent only those we know will change, then we will likely oversample or undersample the genes in one of the samples being compared. If the array contains a large enough assortment of random genes, we do not expect to see such bias. This is because for a finite RNA sample, when representation of one RNA species increases, representation of other species must decrease. Consequently, approximately the same number of labeled molecules from each sample should hybridize to the arrays; therefore, the total hybridization intensities summed over all elements in the arrays should be the same for each sample.

There are many variations for this type of normalization, including scaling the individual intensities so that the mean or median intensities are the same within a single array or across arrays or using a select subset of the arrayed genes rather than the entire collection. One common approach is z-transformation, which not only equalizes the mean of each array but also equalizes the magnitude of variability of individual probe values about that mean. A variant of this approach is to analyze data in the context of an analysis of variance (ANOVA) statistical model that explicitly estimates and removes such differences (Kerr *et al.,* 2000; Kerr and Churchill, 2001). Some investigators also normalize observations of each gene to have the same mean (and standard deviation). This type of normalization is implicit in the use of Pearson correlation coefficients as a "similarity measure" for class discovery analyses. However, for reasons discussed in the following, normalization within genes can actually decrease the signal-to-noise ratio in the dataset as a whole. If gene-specific normalizations are applied, it is critical to ensure that they do not have the effect of equating large biological changes with random noise.

Lowess Normalization

The advantage of Lowess normalization is that it takes systematic biases into consideration. Lowess normalization assumes that the dye bias is dependent on spot intensity. Several reports have indicated that the $\log_2$(ratio) values can have a systematic dependence on intensity (Yang *et al.,* 2002a; Yang *et al.,* 2002b), which most commonly appears as a deviation from zero for low-intensity spots. The easiest way to visualize intensity-dependent effects, and the starting point for the Lowess analysis described here, is to plot the measured $\log_2(R_i/G_i)$ for each element on the array as a function of the $\log_{10}(R_i*G_i)$ product intensities (R-I plot), where G_i and R_i are the measured green and red intensities for the array element in a two-color microarray assay. This "R-I" plot can reveal intensity-specific artifacts in the $\log_2$(ratio) measurements. Locally weighted linear regression (Lowess) analysis has been proposed (Kepler *et al.,* 2002; Yang *et al.,* 2002a; Yang *et al.,* 2002b) as a normalization method that can remove such intensity-dependent effects in the $\log_2$(ratio) values.

Model-Based Normalization

Another popular form of normalization is model-based normalization (model-based expression analysis), which applies a theoretical statistical model to raw fluorescence intensity data in an attempt to infer the true quantity of biological signal (e.g., messenger RNA, or mRNA, concentration). The D-Chip statistical software (Li and Wong, 2003) pursues this approach with high-density oligonucleotide data. Model-based normalization takes into account both global differences across arrays and individual characteristics of the set of probes measuring a specific mRNA target on each array to estimate the true target quantity and remove noise.

All types of normalization are aimed at increasing the signal-to-noise ratio by removing extraneous sources of variance such as instrumentation noise or systematic individual differences across sample donors in biological characteristics. However, no normalization is guaranteed to succeed in this mission, and some methods can actually decrease the signal-to-noise ratio if applied recklessly. It is easy to determine whether normalization has been helpful. If the normalized data show greater consistency across replicates than do the original expression values, normalization has achieved its goals. To determine whether consistency has increased, one can simply compute a correlation between replicate measures for the same experimental condition. In differential expression studies (discussed later), analysis seeks to identify *differences* in gene expression across conditions, and a successful normalization would increase the correlation between measures of that difference across replicate observations. Standardizing data within genes is an example of a normalization that can sometimes decrease the signal-to-noise ratio in the dataset as a whole because this method can equate the magnitude of variation in genes showing only small amounts of absolute variability (likely resulting from noise) with those showing very significant biological changes. Another problem that can arise in normalization is the loss of resolution, particularly at the low end of the expression spectrum in which the majority of genes reside. Excessive normalization can prevent statistical analyses from discovering changes even when they are clearly present in the original data and independently verifiable by nonmicroarray technologies (Cole *et al.,* 2003). Thus all normalizations should be reviewed to ensure that they have succeeded in increasing the signal-to-noise ratio.

Array Element Filtering

If one examines several representative R-I plots, it becomes obvious that the variability in the measured $\log_2$(ratio) values increases as the measured hybridization intensity decreases. This is not surprising because relative error increases at lower intensities, where the signal approaches background. One solution is to use only array elements with intensities with a statistically significant difference from the background. If we measure the average local background near each array

element and its standard deviation, we would expect 95.5% confidence that good elements would have intensities more than two standard deviations above the background. By keeping only array elements that are significantly above the background, we can increase the reliability of measurements. However, this approach may sometimes be too conservative and can needlessly exclude data from analysis. Another solution is to directly estimate the lower bound of reliability in a data set and then raise all values below that level up to this new "floor" value. This approach suppresses noise without excluding data. Software is now available for finding reliability floors (Cole *et al.,* 2003).

A different problem can occur at the high end of the intensity spectrum, where the array elements saturate the detector used to measure fluorescence intensity. Once the intensity approaches its maximum value, comparisons are no longer meaningful because the array elements become saturated and intensity measurements cannot go higher. There are a variety of approaches to deal with this problem as well, including eliminating saturated pixels in the image-processing step or setting a maximum acceptable value ("ceiling") for each array element. Again, software is available for estimating ceiling values based on observed data (Cole *et al.,* 2003).

Classification and Exploratory Data Analysis

The classification and exploratory data analysis is the most challenging step for extracting reliable and interpretable biological information from tens of thousands of microarray data points. Many different statistical approaches have been proposed for carrying out this critical step, but no single approach is appropriate for all purposes. Which strategy is most effective depends heavily on the goals of analysis.

The most common study objectives can be subgrouped into three categories: class comparison, class prediction, and class discovery (Simon *et al.,* 2003). Class comparison, or differential expression analysis, is the comparison of gene expression in different groups of samples. The major characteristic of class comparison studies is that the classes being compared are defined independently of the expression profiles. The specific objectives of such a study are to determine whether the expression profiles are different between the classes and, if so, to identify the differentially expressed genes. One example of a class comparison study is the comparison between gene expression profiles in breast cancer patients with and without germline BRCA1 mutations (Hedenfalk *et al.,* 2001).

Class prediction studies are similar to class comparison studies in that the classes are predefined. However, in class prediction studies, the emphasis is on generating an expression-based multivariate function (referred to as the predictor) that accurately predicts the class membership of a new sample on the basis of the expression levels of key genes. Such predictors can be used for guiding many clinical decisions, such as risk assessment, diagnostic testing, prognostic stratification, and treatment selection. Many studies include both class comparison and class prediction objectives.

Class discovery is fundamentally different from class comparison or class prediction in that no classes are predefined. Usually the purpose of class discovery in cancer studies is to determine whether discrete subsets of a disease entity can be defined on the basis of gene expression profiles. This purpose is different from determining whether the gene expression profiles correlate with some already-known diagnostic classification. Examples of class discovery are the studies by Bittner *et al*. (Bittner *et al.,* 2000) that examined gene expression profiles for advanced melanomas and by Alizadeh *et al*. (Alizadeh *et al.,* 2000) that examined the gene expression profiles of patients with diffuse large B-cell lymphoma. Often the purpose of class discovery is to identify clues regarding the heterogeneity of disease pathogenesis.

Statistical methods will be described here for all three objective categories, class comparison, class prediction, and class discovery.

Class Comparison

One approach for class comparison commonly used in the early microarray studies was a simple fold difference approach, in which a gene is declared to have changed significantly if its average expression level varies by more than a constant factor, typically twofold, between different groups of biological samples (Schena *et al.,* 1995). This approach was replaced quickly with more sophisticated approaches simply because such a "twofold rule" is unlikely to yield optimal results. A factor of 2 can have quite different significance and meaning in different regions of the spectrum of expression levels, especially at the very high and very low ends. In a noisy environment, 2000/1000 or 2/1 can have a quite different significance. Small random fluctuations are much more likely to produce a change from 1 to 2 than from 1000 to 2000.

Another method of identifying differentially expressed genes is to use a t-test, a test for a difference between two means based on t distribution. The significance analysis of microarrays (SAM) is one such implementation

using a t-test (Tusher *et al.,* 2001) (www-stat.stanford.edu/~tibs/SAM/index.html). SAM identifies genes with statistically significant changes in expression by assimilating a set of gene-specific t-tests. Each gene is assigned a score on the basis of its change in gene expression relative to the standard deviation of repeated measurements for that gene. Genes with scores greater than a threshold are deemed potentially significant. SAM uses permutations of the repeated measurements to estimate the percentage of genes identified by chance, the false discovery rate (FDR). The threshold can be adjusted to identify smaller or larger sets of genes, and FDRs are calculated for each set. As demonstrated by the study of Tusher *et al.* (2001) SAM outperforms the fold difference–based approach for identifying genes with true changes in expression (Figure 1).

One potential problem with the t-test for array data, however, is that the repetition numbers are often small because experiments remain costly or tedious to repeat. It is difficult to obtain accurate estimates of the standard deviation of individual gene measurements based on only a few measurements. However, it has been observed that an overall reciprocal relationship exists between variance and gene expression levels and that genes expressed at similar levels exhibit similar variance (Hatfield *et al.,* 2003). Therefore, it is possible to use this prior knowledge in a Bayesian statistical framework to obtain more robust estimates of variance for any gene by examining the expression levels of other genes in the same expression neighborhood within a single experiment (Baldi and Long, 2001; Long *et al.,* 2001). This approach supplements the weak empirical estimates of single-gene variances across a small number of replicates, with more robust estimates of variance obtained by pooling genes with similar expression levels. The CyberT program is an example of this regulated t-test implementation (www.igb.uci.edu). The alternative is to use a strict frequentist approach, in which the estimate of the standard deviation is compromised by the limited number of measurements of each gene that is typical of DNA microarray experiments.

All approaches based on the t-test assume that the investigator is seeking genes with a large signal-to-noise ratio, in the sense that the magnitude of change between conditions is large relative to the magnitude of variability among observations from the same conditions (this is exactly the quantity measured by a t-test statistic). However, many biologists are not so concerned with signal-to-noise ratios *per se* as with the likelihood that a specified change will be reproducible in future experiments. For example, consider a gene showing a 3-fold, 30-fold, and 300-fold increase across three replicated comparisons of healthy and tumor tissue.

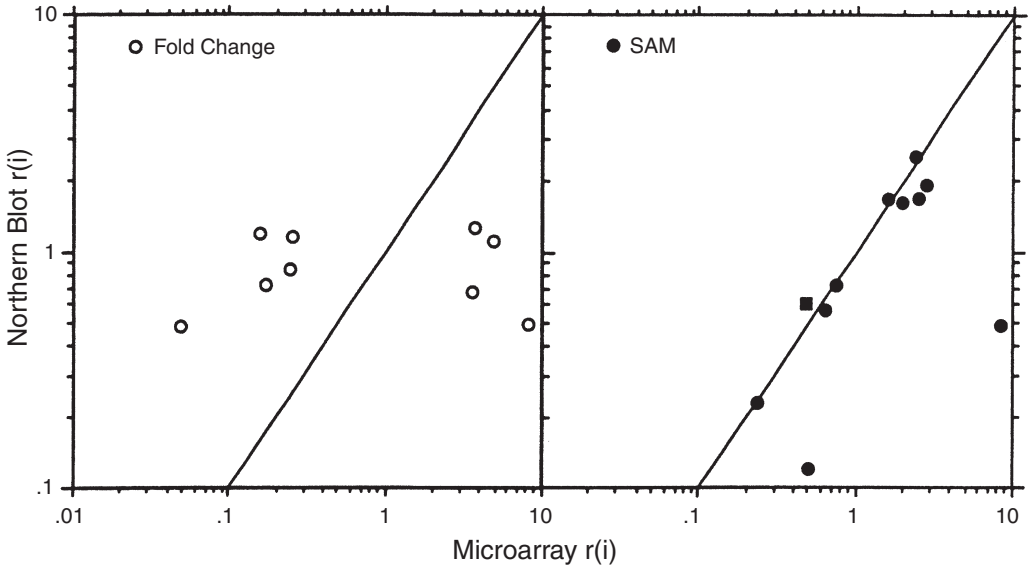

Figure 1. Performance comparison of fold difference approach and t-test–based approach (SAM) for class comparison. Genes showing significant changes between treated and untreated samples were identified by either fold change method or SAM (significance analysis of microarrays). Northern Blots were performed on those genes to validate the fold changes for the identified genes. The expression ratio of treated and untreated samples for each gene was calculated separately using Northern Blot data and microarray data. These ratios were plotted for genes identified by the fold change method (*left*) and SAM (*right*). The straight lines indicate the position of exact agreement between Northern Blot and microarray results. From Tusher *et al.,* 2001.

Most biologists would be very interested in this gene because it consistently shows a large change in expression. However, the magnitude of variability in the change (noise) is very large, and a t-test would not identify this gene as showing statistically significant change (because the average signal is large, but the noise is even larger, so the signal-to-noise ratio is comparatively small). Many microarray data sets do in fact show this sort of qualitiative consistency in the absence of quantitative consistency. To capture qulitatively consistency change that show high variability across samples, one approach is to estimate the lower bound of expected change rather than the mean change (as estimated by the t-test). Instead of using signal-to-noise ratios to distinguish reliable changes from noise, one implementation of this approach used raw fluroscence intensity changes to determine which fold-change results are trustworthy (Cole *et al.,* 2003). This approach captures differentially expressed genes more efficiently than do other methods based on the t-test and related signal-to-noise measures, and it thus reduces the likelihood that biologically significant changes in expression will be overlooked by statistical analyses.

Class Prediction

Class prediction refers to the assignment of particular tumor samples to already-defined classes, which could reflect current states or potential future outcomes. Because it is likely that these predictions will affect clinical decisionmaking, these studies must be performed with statistical rigor and reported clearly with unbiased statistics. One of the most common errors in microarray analysis is the use of cluster analysis and simple fold-change statistics for class prediction and class comparison (Simon *et al.,* 2003). Although cluster analysis is appropriate for class discovery, it is generally not effective for class comparison or class prediction. Cluster analysis is considered an unsupervised method because no information about sample grouping is used. The distance measures are generally computed with regard to the complete set of genes represented on the array that are measured with sufficiently high signals or with regard to all genes that show meaningful variation across the sample set. Because relatively few genes may distinguish any particular class, the distances used in cluster analysis will often not reflect the influence of these relevant genes. This feature accounts for the poor results often obtained in attempting to use cluster analysis for class prediction studies. It is more appropriate to use a supervised method (i.e., one that makes distinctions among the specimens on the basis of predefined class label information) than an unsupervised method, such as cluster analysis (Simon *et al.,* 2003). Supervised class prediction is usually based on the assumption that a collection of differentially expressed genes is associated with class distinction.

The first step toward constructing the class predictor (or classifier) is to select the subset of informative genes. Various class comparison methods described in the previous section can be used here. Once the informative genes are defined, the relative weights correlating to the individual predictive strengths will be assigned to these informative genes. Many methods have been developed for defining a multivariate predictor (Dudoit *et al.,* 2002). A commonly used strategy is based on linear combinations of the weighted intensity measurements of the informative genes (Golub *et al.,* 1999; Radmacher *et al.,* 2002). This procedure uses a fixed set of "informative genes" and makes a prediction on the basis of the expression level of these genes in a new sample. For example, in the study by Golub *et al.,* this weighted votes approach was used to predict patients with acute myeloid leukemia (AML) or acute lymphoblastic leukemia (ALL) (Golub *et al.,* 1999) (Figure 2). Each informative gene casts a "weighted vote" for one of the classes, with the magnitude of each vote dependent on the expression level in the new sample and the degree of that gene's correlation with the class distinction. The votes are summed to determine the winning class, as well as "prediction strength" (PS), which is a measure of the margin of victory that ranges from 0 to 1. The sample is assigned to the winning class if PS exceeded a predetermined threshold and is otherwise considered uncertain.

One alternative method to assign weights related to the individual predictive strengths of these informative genes is to use a dimension reduction technique such as principal components analysis or partial least squares on the informative genes and to base the prediction on the resulting factors (Khan *et al.,* 2001; Nguyen and Rocke, 2002; West *et al.,* 2001). In a careful comparison of alternative class prediction tools, Dudiot and colleagues demonstrated that relatively simple strategies such as linear discriminant function analysis (linear regression) or diagonal discriminant function analysis (similar to the weighted voting scheme outlined earlier) generally outperform more complex and intensive statistical techniques (Dudoit *et al.,* 2002).

Validation of the prediction model is essential because the prediction model generated by supervised methods is invariably more accurate for the data set used to generate the model than it would be with new data. Methods for obtaining unbiased estimates of a

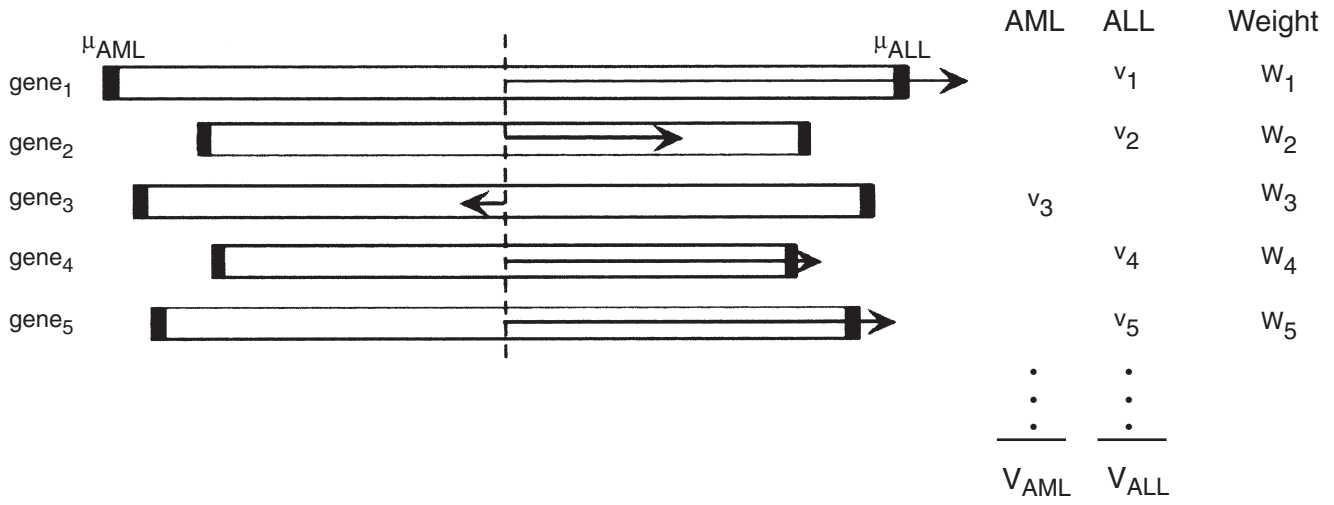

Figure 2. Prediction of acute myeloid leukemia (AML) or acute lymphoblastic leukemia (ALL) using "weighted votes." A set of informative genes is first identified that are more highly correlated with the AML–ALL class distinction than would be expected by chance. The prediction of a new sample is based on "weighted votes" of this set of informative genes. Each gene (g_i) votes for either AML or ALL, depending on whether its expression level x_i in the sample is closer to μ_{AML} or μ_{ALL} (which denote, respectively, the mean expression levels of AML and ALL in a set of reference samples). The magnitude of the vote is $w_i v_i$, where w_i is a weighting factor that reflects how well the gene is correlated with the class distinction and $v_i = |X_i - (\mu_{AML} + \mu_{ALL})/2|$ reflects the deviation of the expression level in the sample from the average of μ_{AML} and μ_{ALL}. The votes for each class are summed to obtain total votes V_{AML} and V_{ALL}. The sample is assigned to the class with the higher vote total, provided that the prediction strength exceeds a predetermined threshold. The prediction strength reflects the margin of victory and is defined as $(V_{win} - V_{lose})/(V_{win} + V_{lose})$, where V_{win} and V_{lose} are the respective vote totals for the winning and losing classes. From Golub *et al.,* 1999.

predictor's error rate include leave-one-out cross-validation or application to an independent data set. By using these techniques, it is possible not only to evaluate the possibility of overfitting the predictor but also to compare various prediction methods and to assess which are less prone to overfitting.

Class Discovery

A general question facing researchers is how to organize observed data into a meaningful structure. For example, biologists need to organize the different species of animals before a meaningful description of the differences between animals is possible. According to the modern system used in biology, humans belong successively to the primates, the mammals, the amniotes, the vertebrates, and the animals. In this classification, the higher the levels of aggregation, the less similar the members in the respective class. Humans have more in common with all other primates (e.g., apes) than they do with the more "distant" members of the mammals (e.g., dogs), etc. This classification structure is a cluster, which can be applied for class discovery. For the purpose of class discovery based on microarray data, clustering is defined as grouping together objects (genes or samples) with similar properties (Eisen *et al.,* 1998). This can also be viewed as the reduction of the dimensionality of the system. Clustering is not a new technique; many algorithms have been developed for it, and many of these algorithms have been applied to microarray studies for class discovery. Examples of using clustering techniques for class discovery include the studies by Bittner *et al.* (2000) that examined gene expression profiles for advanced melanomas and by Alizadeh *et al.* (2000) that examined the gene expression profiles of patients with diffuse large B-cell lymphoma. The two major styles of clustering algorithms, hierarchical and *k*-means clustering, and self-organizing maps have all been used for clustering expression profiles.

Hierarchical Clustering

Hierarchical clustering is an agglomerative approach in which the closest elements in the data set are joined to form groups, which are further joined until the process has been carried to completion, forming a single hierarchical tree. This method is also referred to as joining or tree clustering. Once the tree is constructed, the data can be partitioned into any number of clusters by cutting the tree at the appropriate level.

The process of hierarchical clustering proceeds in a simple manner. First, the pairwise distance matrix is calculated for all genes to be clustered. Second, the

distance matrix is searched for the two most similar genes or clusters; initially each cluster consists of a single gene. This is the first true stage in the "clustering" process. If several pairs have the same separation distance, a predetermined rule is used to decide between alternatives. Third, the two selected clusters are merged to produce a new cluster that now contains at least two objects. Fourth, the distances are calculated between this new cluster and all other clusters. There is no need to calculate all distances because only those involving the new cluster have changed. Last, steps 2–4 are repeated until all objects are in one cluster.

Three common options for hierarchical clustering are single linkage, average linkage, and complete linkage. These options differ in their definition of the distance between two clusters. Single linkage defines the distance between clusters as the minimum distance over all pairs. Average linkage takes the average distance over all pairs, and complete linkage uses the maximum distance over all pairs. Each of these options will produce slightly different results, as will any of the algorithms if the distance metric is changed. Single linkage often produces large, elongated clusters (Butte and Kohane, 2000). Complete linkage finds small, compact clusters that do not exceed some diameter threshold. The threshold value is determined by the level at which the tree is cut. Average linkage is sometimes used as a compromise between the other two options and gives acceptable results typically for gene-expression data (Eisen *et al.,* 1998). Figure 3 shows an example of hierarchical clustering on 31 melanoma samples. This hierarchical dendrogram demonstrates that 19 samples are tightly clustered at the bottom of the dendrogram in the area of highest similarity, which may represent a subset of melanoma (Bittner *et al.,* 2000).

One alterative type of hierarchical clustering, the two-way clustering, is particularly interesting. The hierarchical clustering usually focuses on cases (patients) or variables (genes). It turns out that the clustering of both may yield useful results. For example, imagine a study in which a researcher has generated microarray data on differential expressed genes (variables) for a group of patients with cancer (cases). The researcher may want to cluster cases (patients) to detect clusters of patients with similar syndromes. At the same time, the researcher may want to cluster variables (genes) to detect clusters of measures that appear to tap similar physical abilities. A good example of using this two-way clustering is illustrated in the study on tumor and normal colon tissues by Alon *et al*. (1999).

Hierarchical clustering has become one of the most widely used techniques for class discovery based on gene-expression data because it is a simple process and the result can be easily visualized (Eisen and Brown, 1999). The first report by Wen *et al*. (1998) uses clustering and data-mining techniques to analyze large-scale gene-expression data. This report is significant in that it shows how integrating results obtained by using various distance metrics can reveal different but meaningful patterns in the data. Eisen *et al*. (1998) also make an elegant demonstration of the power of hierarchical clustering in the analysis of microarray data.

However, there are several problems with the ability to adequately describe the data. The decision to include a pair of genes in the same cluster is based only on a specific distance between them (introduced to quantitatively characterize the degree of coregulation), and any such decision is final. The local nature of this decision rule often inhibits the algorithm's ability to find a global structure. Also, the hierarchical trees are more suited for the description of real hierarchical relationships (e.g., evolutionary processes), although there is no evidence for the existence of such relationships in the biological functions of different genes. The expression vector that represents the cluster might no longer represent any of the genes in the cluster as the clusters grow in size. Consequently, as clustering progresses, the actual expression patterns of the genes themselves become less relevant. In addition, if a bad assignment is made early in the process, it cannot be corrected. Furthermore, the resulting structure is complex and there is no general agreement on how to choose the location for cutting the tree.

An alternative, which avoids these artifacts, is to use a divisive clustering approach, such as *k*-means or self-organizing maps, to partition data (either genes or experiments) into groups that have similar expression patterns.

k-Means Clustering

If there is advance knowledge about the number of clusters that should be represented in the data, *k*-means clustering is a good alternative to hierarchical methods. The *k*-means algorithm (Herzel *et al.,* 2001; Tavazoie *et al.,* 1999) is a nonhierarchical clustering technique (a divisive clustering approach). In this clustering technique, objects are partitioned into a fixed number (k) of clusters, so that the clusters are internally similar but externally dissimilar.

The *k*-mean clustering algorithm is a conceptually simple three-step process but can be computationally intensive. In the first step, the algorithm randomly assigns all training data to one of k clusters (where k is specified by the user). In the second step, the mean intercluster and intracluster distances are calculated. The third step is an iterative step, and its goal is to minimize the mean intercluster distances, maximize intracluster distances, or both, by moving data from

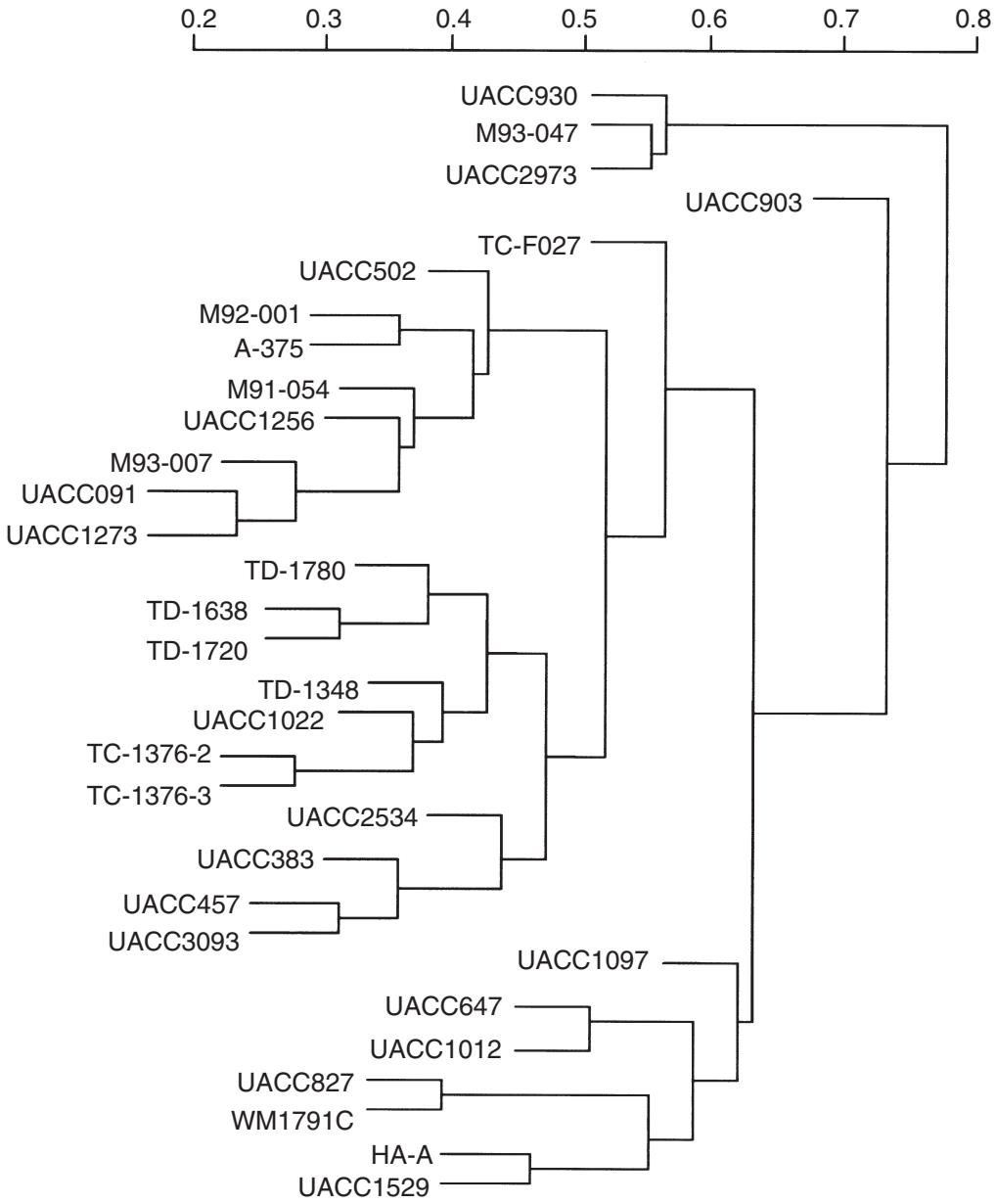

Figure 3. Hierarchic clustering dendrogram on 31 cutaneous melanoma samples based on overall expression pattern measured by microarray analysis. From Bittner *et al.,* 2000.

one cluster to another. In each iteration, one piece of data is moved to a new cluster where it is closest to the mean vector of the new cluster. After each move, new expression vectors for the two affected classes are recalculated. This process continues until any further move would increase the mean intercluster means (expression variability for each cluster) or reduce intracluster distances.

An advantage of the *k*-means algorithm is that, because of its simplicity, it can be used in a variety of applications. For instance, a recent variant of the *k*-means clustering algorithm designed specifically for the assessment of gene spots (on the array images) is the work of Bozinov and Rahnenfuhrer (2002). This technique is based on clustering pixels of a target area into foreground and background clusters. *k*-means clustering is particularly useful with techniques such as principal component analysis (PCA). PCA allows a visual estimation of the number of clusters represented in the data. This can be used to specify *k* and to group genes (or experiments) into related clusters.

The major disadvantage of this method is that the number *k* is often not known in advance. Another potential problem with this method is that, because each

gene is uniquely assigned to some cluster, it is difficult for the method to accommodate a large number of stray data points, intermediates, or outliers. Additional concerns about the algorithm have to do with the biological interpretation of the final clustered data from the algorithm. Despite these difficulties, *k*-means clustering results often show more stability across repeated experiments than do other class discovery tools.

Self-Organizing Maps

A self-organizing map (SOM) is a neural-network–based divisive clustering approach (Kohonen, 2001). Neural networks are analytic techniques modeled after the processes of learning in cognitive systems and the neurologic functions of the brain. Neural networks use a data "training set" to build rules capable of making predictions or classifications on data sets. A SOM assigns genes to a series of partitions on the basis of the similarity of their expression vectors to reference vectors that are defined for each partition. It is the process of defining these reference vectors that distinguishes SOMs from *k*-means clustering. Before initiating the analysis, the user defines a geometric configuration for the partitions, typically a two-dimensional rectangular or hexagonal grid. Random vectors are generated for each partition, but before genes can be assigned to partitions, the vectors are first "trained" using an iterative process that continues until convergence occurs so that the data are most effectively separated.

First, random vectors are constructed and assigned to each partition. Second, a gene is chosen at random, and, using a selected distance metric, the reference vector that is closest to that gene is identified. Third, the reference vector is then adjusted so that it is more similar to the vector of the assigned gene. The reference vectors that are located nearby on the two-dimensional grid are also adjusted so that they are more similar to the vector of the assigned gene. Fourth, steps 2 and 3 are iterated several thousand times, decreasing the amount by which the reference vectors are adjusted and increasing the stringency used to define closeness in each step. As the process continues, the reference vectors converge to fixed values. Finally, the genes are mapped to the relevant partitions depending on the reference vector to which they are most similar.

Several groups have used SOMs to investigate patterns in gene expression data (Golub *et al.,* 1999; Tamayo *et al.,* 1999; Toronen *et al.,* 1999). SOMs have the distinct advantage that they allow *a priori* knowledge to be included in the clustering process. In choosing the geometric configuration for the clusters, the user is effectively specifying the number of partitions into which the data is to be divided. Tamayo *et al.* propose that SOMs are ideally suited for exploratory data analysis, allowing one to impose partial structure on the clusters (in contrast to the rigid structure of hierarchical clustering, the strong prior hypotheses used in Bayesian clustering, and the nonstructure of *k*-means clustering) and facilitating easy visualization and interpretation (Tamayo *et al.,* 1999).

The most prominent disadvantage of the SOM approach is that it is difficult to know when to stop the algorithm. If the map is allowed to grow indefinitely, the size of the SOM is gradually increased to a point where clearly different sets of expression patterns are identified. Therefore, as with *k*-means clustering, the user has to rely on some other source of information, such as PCA, to determine the number of clusters that best represents the available data. For this reason, Sasik and his colleagues believe that SOM, as implemented by Tamayo *et al.* (1999), is essentially a restricted version of *k*-means (Sasik *et al.,* 2001).

Underlying all the class discovery techniques presented here is the problem of the "basis" for clustering. Class discovery tools group genes (or cases) based on a similarity score quantifying the consistency between each gene's profile of expression across cases (or each case's profile of expression across genes) and that profile of expression for all other genes (or cases). The definition of this similarity score profoundly affects the eventual "solution" derived by class discovery algorithms, but there is no clear consensus on which similarity metric is most appropriate. Broadly speaking, there are two basic possibilities. One is a "scalefree" representation that standardizes the mean and variance of data from each gene (or case) before comparing that scalefree profile with all other scalefree profiles. In many class discovery programs, a Pearson or Separman correlation coefficient is used by default as the similarity metric, and this implicitly invokes the scalefree approach. Under this approach, random noise can be inflated to a magnitude comparable to significant biological changes unless careful noise suppression is applied before analysis. Because microarrays survey so many genes, although comparatively few are expressed in any given sample, there is ample opportunity for noise to generate spurious results that subsequently detract from the replicability of results. The alternative is to retain information about the variance of observations on each gene, in which case noise influences are effectively reduced, but the eventual class discovery solution is dominated by the distinction between highly expressed genes and those expressed at low levels. Most researchers have used the scalefree approach, but this approach will produce reliable classes only if noise is fully suppressed before analysis (e.g., by floor/ceiling

analyses). Even then, general class definitions may be stable, but the specific genes included in each class can vary substantially across repeated experiments.

Postanalysis Follow-Up and Validation of Microarray Data

Validation is essential for microarray-based experiments. It has been proposed that at least two important questions need to be considered when evaluating microarray data (Chuaqui *et al.,* 2002). First, is the microarray data valid? Second, do the data or the model systems based on those data fundamentally describe the phenomenon being investigated? The postanalysis validation will serve as the quality-check step and provide feedback information for refining the model system generated based on microarray data.

Validation of the Experimental Data

Validation of the microarray expression data is essential. Some researchers have considered performing the data validation step even before advanced data analysis (such as classification and exploratory data analysis), so that only the validated data will be used for classification. However, in most microarray studies, validation before classification involves a lot more genes than does validation after the classification. This is because the classification step is a data reduction step that only keeps the highly informative genes that contribute to the classification. There are two approaches for independent confirmation of microarray data: *in silico* analysis and laboratory-based analysis. The *in silico* approach has been reviewed previously; it compares array results with information available in the literature and in public or private expression databases and provides the opportunity to validate data without further experimentation (Chuaqui *et al.,* 2002).

Laboratory-based validation provides independent, experimental verification of gene-expression levels and typically begins with the same samples that were studied in the microarray experiment. The methodology used varies depending on the scientific question, but commonly used techniques include quantitative reverse transcription (RT)–PCR, Northern Blot, ribonuclease protection assay, and *in situ* hybridization or immunohistochemistry using tissue microarrays. Quantitative RT-PCR is the choice of many for quantitatively measuring specific mRNAs because, once established, the method is fast, accurate, reproducible, and relatively inexpensive and requires a minimal amount of starting template (Rajeevan *et al.,* 2001; Walker, 2002).

Validation of the Model Generated from Microarray Data

Many gene-expression profiling studies in human cancers attempt to build predictors (prediction models) of patient prognosis and response to therapy (Rosenwald *et al.,* 2002; Shipp *et al.,* 2002). Because it is likely these predictors will provide information that could affect clinical decisionmaking, such studies must be performed with statistical rigor and be reported clearly and with unbiased statistics. All predictors should be validated. If cross-validation is used to estimate prediction accuracy, then the entire model-building process, including the selection of informative genes, should be repeated in each cross-validation training set. If a separate sample group is used for validation, the sample size should be sufficiently large to provide meaningful confidence intervals for prediction accuracy.

Looking Ahead

The use of microarray and other global-profiling technologies has led to a significant number of exciting new biological discoveries and important correlations between gene-expression patterns and disease states. Never before could a small sample of RNA from two different conditions reveal so much information at the transcriptional level. Microarray technologies have also provided high-throughput platforms for genomewide SNP genotyping (Lindblad-Toh *et al.,* 2000a; Lindblad-Toh *et al.,* 2000b; Mei *et al.,* 2000) and CGH for detecting DNA copy number abnormality (Albertson and Pinkel, 2003; Pollack *et al.,* 1999). At the protein level, there have also been great advances in expression quantification through micro-ELISA (enzyme-linked immunosorbent assay) protein arrays (Huang *et al.,* 2001; Tam *et al.,* 2002), dual-labeling displays of protein expression on an array (Clontech, Inc.), or protein chips in combination with MALDI (matrix-assisted laser desorption/ionization) mass spectrometry for protein identification (Ciphergen, Inc.). Contemporary molecular biologists are endowed with powerful new tools to explore the genome and proteome of many species for both basic biological analyses and the discovery of new disease influences or drug targets.

Nonetheless, it is important that investigators continue to optimize statistical methodologies and to develop new approaches to produce accurate and valid data. In particular, much more attention needs to be paid to the external validity of the results produced by microarray data analysis tools. When a statistical tool identifies a difference, or a cluster of related genes, how often does this indication hold up in subsequent

validation studies? Most current statistical approaches to microarray data have only been presented with examples of application, but no quantitative evaluation of their accuracy with regard to external verification is available. Alternative class prediction tools have been compared in terms of their predictive accuracy, and similar comparisons are sorely needed for class discovery and differential expression analyses. Another issue requiring further study is the balance between false-positive and false-negative errors. External validation studies have suggested that many differential expression analyses of microarray data overlook true differences even when they are present in the data (false-negative results). An important topic for future research is whether development of more sensitive, noise-resistant algorithms might enhance the information yield from massively parallel microarray measurement. We probably are still a long way from a satisfactory solution to the problems. With its potential to quantitatively determine expression levels of a large number of genes in parallel, microarray technology holds the promise of becoming an extremely valuable tool in basic biological sciences and clinical diagnostics. However, its ultimate utility will depend critically on whether the search for efficient statistical methods meets with success.

References

Albertson, D.G., and Pinkel, D. 2003. Genomic microarrays in human genetic disease and cancer. *Hum. Mol Genet. 12 Suppl 2*:R145–152.

Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., Powell, J.I., Yang, L., Marti, G.E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D.B., Tibshirani, R., Sherlock, G., Chan, W.C., Greiner, T.C., Weisenburger, D.D., Armitage, J.O., Warnke, R., Levy, R., Wilson W., Grever, M.R., Byrd, J.C., Botstein, D., Brown, P.O., and Staudt, L.M. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature 403*(6769):503–511.

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., and Levine, A.J. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA 96*(12):6745–6750.

Baldi, P., and Long, A.D. 2001. A Bayesian framework for the analysis of microarray expression data: Regularized t-test and statistical inferences of gene changes. *Bioinformatics 17*(6):509–519.

Bittner, M., Meltzer, P., Chen, Y, Jiang, Y, Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., and Sondak, V. 2000. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature 406*(6795):536–540.

Bozinov, D., and Rahnenfuhrer, J. 2002. Unsupervised technique for robust target separation and analysis of DNA microarray spots through adaptive pixel clustering. *Bioinformatics 18*(5): 747–756.

Butte, A.J., and Kohane, I.S. 2000. Mutual information relevance networks: Functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.* 418–429.

Cahill, D.J., and Nordhoff, E. 2003. Protein arrays and their role in proteomics. *Adv. Biochem. Eng. Biotechnol. 83*:177–187.

Chatterjee, S., and Price, B. 1991. *Regression Analysis by Example*. John Wiley & Sons, New York.

Chen, Y., Dougherty, M.L., and Bittner, M.L. 1997. Ratio-based decisions and the quantitative analysis of cDNA microarray images. *J. Biomed. Optics 24*:364.

Chen, Y. Kamat, V., Dougherty, E.R., Bittner, M.L., Meltzer, P.S., and Trent, J.M. 2002. Ratio statistics of gene expression levels and applications to microarray data analysis. *Bioinformatics 18*(9):1207–1215.

Chuaqui, R.F., Bonner, R.F., Best, C.J., Gillespie, J.W., Flaig, M.J., Hewitt, S.M., Phillips, J.L., Krizman, D.B., Tangrea, M.A., Ahram, M., Linehan, W.M., Knezevic, V., and Emmert-Buck, M.R. 2002. Post-analysis follow-up and validation of microarray experiments. *Nat. Genet. 32 Suppl*:509–514.

Cole, S.W., Galic, Z., and Zack, J.A. 2003. Controlling false-negative errors in microarray differential expression analysis: A PRIM approach. *Bioinformatics 19*(14):1808–1816.

DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., and Trent, J.M. 1996. Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet. 14*(4):457–460.

Dudoit, S., Fridlyand, J., and Speed, T. 2002. Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc. 97*:77–87.

Eisen, M.B., and Brown, P.O. 1999. DNA arrays for analysis of gene expression. *Methods Enzymol. 303*:179–205.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA 95*(25): 14863–14868.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., and Lander, E.S. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science 286*(5439):531–537.

Hatfield, G.W., Hung, S.P., and Baldi, P. 2003. Differential analysis of DNA microarray gene expression data. *Mol. Microbiol. 47*(4):871–877.

Hedenfalk, I., Duggan, D., Chen, Y, Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.P., Wilfond, B., Borg, A., and Trent, J. 2001. Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med. 344(*8):539–548.

Herzel, H., Beule, D., Kielbasa, S., Korbel, J., Sers, C., Malik, A., Eickhoff, H., Lehrach, H., and Schuchhardt, J. 2001. Extracting information from cDNA arrays. *Chaos 11*(1):98–107.

Huang, J.X., Mehrens, D., Wiese, R., Lee, S., Tam, S.W., Daniel, S., Gilmore, J., Shi, M., and Lashkari, D. 2001. High-throughput genomic and proteomic analysis using microarray technology. *Clin. Chem. 47*(10):1912–1916.

Kepler, T.B., Crosby, L., and Morgan, K.T. 2002. Normalization and analysis of DNA microarray data by self-consistency and local regression. *Genome Biol. 3*(7):RESEARCH0037.

Kerr, M.K., and Churchill, G.A. 2001. Experimental design for gene expression microarrays. *Biostatistics 2*(2):183–201.

Kerr, M.K., Martin, M., and Churchill, G.A. 2000. Analysis of variance for gene expression microarray data. *J. Comput. Biol. 7*(6):819–837.

Khan, J., Wei, J.S., Ringner, M., Saal, L.H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C.R., Peterson, C., and Meltzer, P.S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med. 7*(6): 673–679.

Kohonen, T. 2001. *Self-Organizing Maps*. Springer, Berlin, Heidelberg, New York.

Li, C., and Wong, W.H. 2003. DNA-Chip Analyzer (dChip). In *The Analysis of Gene Expression Data: Methods and Software*. Springer, New York.

Lindblad-Toh, K., Tanenbaum, D.M., Daly, M.J., Winchester, E., Lui, W.O., Villapakkam, A., Stanton, S.E., Larsson, C., Hudson, T.J., Johnson, B.E., Lander, E.S., and Meyerson, M. 2000a. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat. Biotechnol. 18*(9):1001–1005.

Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Laviolette, J.P., Ardlie, K., Reich, D.E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.B., Gingeras, T., Warrington, J., Patil, N., Hudson, T.J., and Lander, E.S. 2000b. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet. 24*(4): 381–386.

Long, A.D., Mangalam, H.J., Chan, B.Y., Tolleri, L., Hatfield, G.W., and Baldi, P. 2001. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in Escherichia coli K12. *J. Biol. Chem. 276*(23): 19937–19944.

Mei, R., Galipeau, P.C., Prass, C., Berno, A., Ghandour, G., Patil, N., Wolff, R.K., Chee, M.S., Reid, B.J., and Lockhart, D.J. 2000. Genome-wide detection of allelic imbalance using human SNPs and high-density DNA arrays. *Genome Res. 10*(8):1126–1137.

Nguyen, D.V., and Rocke, D.M. 2002. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics 18*(1):39–50.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet. 23*(1):41–46.

Quackenbush, J. 2002. Microarray data normalization and transformation. *Nat. Genet. 32 Suppl*:496–501.

Radmacher, M.D., McShane, L.M., and Simon, R. 2002. A paradigm for class prediction using gene expression profiles. *J. Comput. Biol. 9*(3):505–511.

Rajeevan, M.S., Vernon, S.D., Taysavang, N., and Unger, E.R. 2001. Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. *J. Mol. Diagn. 3*(1):26–31.

Rosenwald, A., Wright, G., Chan, W.C., Connors, J.M., Campo, E., Fisher, R.I., Gascoyne, R.D., Muller-Hermelink, H.K., Smeland, E.B., Giltnane, J.M., Hurt, E.M., Zhao, H., Averett, L., Yang, L., Wilson, W.H., Jaffe, E.S., Simon, R., Klausner, R.D., Powell, J., Duffey, P.L., Longo, D.L., Greiner, T.C., Weisenburger, D.D., Sanger, W.G., Dave, B.J., Lynch, J.C., Vose, J., Armitage, J.O., Montserrat, E., Lopez-Guillermo, A., Grogan, T.M., Miller, T.P., LeBlanc, M., Ott, G., Kvaloy, S., Delabie, J., Holte, H., Krajci, P., Stokke, T., and Staudt, L.M. 2002. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med. 346*(25):1937–1947.

Sasik, R., Hwa, T., Iranfar, N., and Loomis, W.F. 2001. Percolation clustering: A novel approach to the clustering of gene expression patterns in Dictyostelium development. *Pac. Svmp. Biocomput.* 335–347.

Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science 270*(5235): 467–470.

Shipp, M.A., Ross, K.N., Tamayo, P., Weng, A.P., Kutok, J.L., Aguiar, R.C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G.S., Ray, T.S., Koval, M.A., Last, K.W., Norton, A., Lister, T.A., Mesirov, J., Neuberg, D.S., Lander, E.S., Aster, J.C., and Golub, T.R. 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat. Med. 8*(1):68–74.

Simon, R., Radmacher, M.D., Dobbin, K., and McShane, L.M. 2003. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J. Natl. Cancer Inst. 95*(1):14–18.

Tam, S.W., Wiese, R., Lee, S., Gilmore, J., and Kumble, K.D. 2002. Simultaneous analysis of eight human Th1/Th2 cytokines using microarrays. *J. Immunol. Methods 261*(1–2):157–165.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub, T.R. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA 96*(6):2907–2912.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. 1999. Systematic determination of genetic network architecture. *Nat. Genet. 22*(3):281–285.

Toronen, P., Kolehmainen, M., Wong, G., and Castren, E. 1999. Analysis of gene expression data using self-organizing maps. *FEBS Lett. 451*(2):142–146.

Tseng, G.C., Oh, M.K., Rohlin, L., Liao, J.C., and Wong, W.H. 2001. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. *Nucleic Acids Res. 29*(12):2549–2557.

Tusher, V.G, Tibshirani, R., and Chu, G 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA 98*(9):5116–5121.

Walker, N. J. (2002). Tech.Sight. A technique whose time has come. *Science 296*(5567):557–559.

Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L., and Somogyi, R. 1998. Large-scale temporal gene expression mapping of central nervous system development. Proc. *Natl. Acad. Sci. USA 95*(1):334–339.

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J.A., Jr., Marks, J.R., and Nevins, J.R. 2001. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci. USA 98*(20):11462–11467.

Yang, l.V., Chen, E., Hasseman, J.P., Liang, W., Frank, B.C., Wang, S., Sharov, V., Saeed, A.I., White, J., Li, J., Lee, N.H., Yeatman, T.J., and Quackenbush, J. 2002a. Within the fold: Assessing differential expression measures and reproducibility in microarray assays. *Genome Biol. 3*(11):research0062.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. 2002b. Normalization for cDNA microarray data: A robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res. 30*(4):e15.