[Frontiers in Bioscience 7, c74-82, August 1, 2002]

EXPRESSION MICROARRAY ANALYSIS OF BRAIN TUMORS: WHAT HAVE WE LEARNED SO FAR

Stephen B. Hunter, Carlos S. Moreno

Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, Georgia

TABLE OF CONTENTS

1. ABSTRACT

This review covers the emerging field of expression microarray technology as applied to human brain tumors. Dual and single color techniques are described and contrasted, and the importance of proper handling of the starting material is emphasized. Difficulties with data interpretation are described and current approaches to cluster analysis reviewed. Microarray studies of general importance or specifically pertaining to brain tumors, published in the initial few years of this technology, are summarized. Although this technology is still in its infancy, microarray has distinguished prognostic groups within medulloblastomas and separated medulloblastomas from morphologically identical supratentorial PNETs. Differential expression of a number of genes previously known to be involved in the pathogenesis of brain tumors has been confirmed. These genes include EGFR, VEGF, transcription factor AP-2, insulin growth factor binding proteins 3 and 5, matrix metalloproteinases, tissue inhibitors of matrix metalloproteinases, CD44, basic fibroblast growth factor, and cathepsin H. Finally, novel roles for a few genes, including insulin growth factor binding protein 2 and apolipoprotein D, have been revealed for the first time by expression microarrays.

2. INTRODUCTION

Cancer is a process in which genes and gene expression are disrupted causing abnormal cell behavior. Some of the characteristic changes observed as normal cells transform into cancer cells include self-sufficiency without growth signals, insensitivity to anti-growth signals, inhibition of apoptosis, unlimited potential for cell division, increased angiogenesis, and increased cellular motility (1). Most of these changes are reflected in corresponding differences in mRNA transcription patterns. The potential of microarray technology to quantitatively assay the mRNA levels of essentially every human gene in a single experiment has naturally kindled high hopes for elucidating the molecular mechanisms of oncogenesis.

Although still in its infancy, current microarray expression techniques have highlighted individual genes involved in malignant behavior and have shown the potential to significantly improve prognostic prediction in selected human malignancies. To what extent this technology will ultimately contribute to cancer research and improvements in cancer therapies remains to be seen.

This review will summarize contributions that the emerging field of expression microarray technology has made to the study of human brain tumors up to the present time. It will focus on a few genes that initial studies have implicated in the pathogenesis of human brain tumors, including insulin growth factor binding protein 2, and apolipoprotein D. It will also cover some of the potentials and pit falls of microarray assays.

Although beyond the scope of this review, there is a pressing need for standardization of microarray data formats among investigators. The microarray gene expression data group (http://www.mged.org) is currently working to address this issue and has proposed, as an academic standard, the minimum information about a microarray experiment (MIAME)(2).

3. SINGLE- vs. TWO-COLOR PLATFORMS

Microarray assays are performed by specific hybridization to cDNA or oligonucleotide probes immobilized on a glass or silicon substrate. The main strength of microarray lies in its high throughput, i.e., the capacity to assay tens of thousands of different mRNA species simultaneously. Compared with subtractive hybridization or differential display, microarray technology is more rapid and quantitative, and it does not require the large amounts of sequencing necessary for SAGE analysis. However, since false positive and negative results occur, data needs to be confirmed by conventional methods such as quantitative real-time PCR, Northern blots, Western blots, or immunohistochemistry.

Two main types of microarray systems are single color (sold by Affymetrix) and dual color. The most recent set of Affymetrix human GeneChips, the HG-U133 Array set, assays 33,000 genes, essentially the entire human genome, on two microarrays, each approximately the size of a large postage stamp. In this system, short oligonucleotide probes are used allowing larger numbers of probes to be affixed per microarray. Each gene is represented by a set of eleven or more probe pairs that are 25 base pairs in length and correspond to different regions of the gene. Each probe pair consists of one probe exactly matching the base pair sequence of the target gene, and one probe with a mismatch in the central position. Probes are selected to minimize cross-reactions due to alternatively spliced or homologous genes, and then synthesized *in situ* on a silicon chip by a proprietary photolithographic procedure. Gene expression is calculated from the brightness and consistency of hybridization to the multiple different probes representing each gene using Affymetrix software. Although reference RNA is not required for single color analysis, data is frequently compared to a reference RNA assayed on a separate microarray.

In contrast with the Affymetrix single color approach, dual color systems measure the relative mRNA abundance of two samples that simultaneously bind to cDNA or oligonucleotide probes spotted on a glass slide. The experimental sample is typically labeled with the fluorescent ribonucleotide Cy5, while the reference sample is labeled with Cy3. The two labeled populations of cDNA or cRNA are then simultaneously hybridized to the microarray, fluorescent intensity is quantitated by a confocal scanning laser, and the difference in brightness at the two wavelengths is used to calculate the differential expression. The data in both single and dual color systems is normalized against the overall measured brightness of the fluorescent markers across the entire microarray or across selected house keeping genes. The result is an expression profile, i.e., a list of differentially over or under expressed genes relative to the reference mRNA.

One of the main technical problems of microarray assays is the variability in hybridization kinetics inherent in assaying thousands of probes simultaneously. In the Affymetrix single color system, variability in hybridization kinetics is circumvented by assaying each gene with several oligonucleotide probes, half containing a centrally located mismatched base pair. In this system, experimental data can be compared on the computer with any number of normal, or control hybridizations performed on the same type of micoarray plate. In dual color systems, this problem is circumvented by measuring mRNA hybridization relative to a reference mRNA sample. Thus, in dual color systems all comparisons require a single common reference mRNA. Data from experiments performed on one color microarrays can be compared to data from two color microarrays qualitatively, (i.e., up or down regulated), but not quantitatively.

4. SPECIMEN PREPARATION

Microarray analysis requires 5 to 75 micrograms of total RNA, depending on the particular system. This corresponds to a minimum of approximately 300 nanograms of mRNA or 30,000 cells. Fresh or snap frozen tissues are used, and tissue is selected to contain as pure a population of cells as possible.

One of the major challenges in generating microarray data is to isolate RNA from a specific cell population without cross-contamination from normal or inflammatory cells included within the specimen. In most cases this can be accomplished by Laser-capture microdissection (LCM). While LCM offers the advantage of a relatively pure tumor cell population, it has the disadvantage that the total RNA yield is generally less than one microgram, considerably less than the amount required for hybridization. Two very similar protocols using recombinant T7 polymerase to linearly amplify LCM-derived mRNA have been used with success for DNA microarray analysis (3, 4). Using this process, microarray analysis has been successfully performed on LCM captures from as few as 1000 cells (5). However, because of the generally lower RNA yield from LCM tissues, most published microarray studies of cancers have not used LCM tissues as starting material (6-8). In the case of infiltrating astrocytomas, particularly low-grade tumors, completely excluding non-neoplastic elements may be virtually impossible because of the close intermingling and difficulty distinguishing invasive neoplastic cells and non-neoplastic elements.

In microarray analyses of brain tumors, normal brain tissue is the most commonly used source of reference mRNA in either single or dual color systems. However, because brain tissue is complex compared with other organs, this tissue is not ideal for comparison in most studies. The most common primary brain tumors are astrocytomas; while normal brain tissue, either gray or white matter, consists of only approximately 20% astrocytes. Consequently, use of normal brain tissue as a reference may disclose cell type expression differences that are unrelated to neoplasia. These differences may be between astrocytes and neurons, when gray matter is used as a reference, or between astrocytes and oligodendrocytes, when white matter is used as a reference. Approximately 80% of the cells in white matter are oligodendrocytes.

In addition, since normal brain tissue is neither proliferative nor motile, it is not surprising that studies using normal brain as a reference have shown that astrocytomas differentially express proliferation and motility related genes. Proliferation and motility are not unique to malignancy; rather, it is loss of control of proliferation and motility that defines malignant behavior. Theoretically, non-neoplastic reactive astrocytes, which are motile and proliferative, might serve as a better source of reference RNA than normal brain; however, pure populations of reactive astrocytes are unfortunately not available. As an alternatively to normal brain, fetal astrocytes(9) or pilocytic astrocytomas(10, 11), a biologically distinct and generally non-fatal type of astrocytoma, have been used as a source of reference mRNA.

5. DATA INTERPRETATION

Perhaps the most formidable challenge facing the microarray investigator is data analysis, i.e. interpreting that long and daunting list of up and down regulated genes, most of which are of unknown function. Several factors contribute to difficulties interpreting microarray data. The function of most genes and their interactions are either not understood or poorly understood. Variability in hybridization kinetics may render comparisons between the expression levels of two different genes difficult. Cells that might show distinct expression patterns, i.e. those in different phases of the cell cycle or belonging to different clones, are lumped together and assayed at the same time.

Another problem that can lead to difficulty interpreting microarray data is the considerable variability in results that sometimes occurs. This variability may be either technical or biological in origin. For example according to Affymetrix, microarray analysis of two supposedly identically treated isogenic mice produces detectable differences in approximately 20% of the genes on the microarray. According to this manufacturer, the source of this variability is overwhelmingly biologic and not technical. However, others have found that biologic and technical factors contribute more equally to variable results (12).

Microarray data may be organized either according to prior knowledge concerning gene function, or statistically without regard to gene function. Some studies have limited the genes assayed to those of known function, or grouped differentially expressed genes into different functional groups based on previous knowledge about gene function. For example, up regulated genes may be grouped into those involved in proliferation, invasion or DNA repair (i.e., progression). These types of analyses ignore genes of unknown function, which comprise the majority of genes in the human genome, and also tend to highlight genes that have already been studied in human malignancies.

Statistical approaches have been used to address the challenge of interpreting variable results and separating true biologic signal from technical noise. In general, the larger the number of patient samples in the database, the easier it is to demonstrate statistical significance and determine which genes are genuinely characteristic of various cancer subtypes.

The first step in most statistical analyses is prefiltering out those genes that do not vary between samples. These genes can be used to estimate the overall noise in the microarray experiment, but are not useful for detecting meaningful differences. Next, subsets of genes or samples highly correlated with each other are grouped into clusters. By clustering together genes that behave similarly across a set of experiments, it may be possible to identify sets of coordinately regulated genes. In some cases, a gene of unknown function may behave similarly to a group of genes of known function, allowing the investigator to infer functional relationships.

Clustering may be supervised or unsupervised. In supervised clustering, or “class prediction,” a subset of genes that is highly correlated with a pre-defined distinction is identified. Pre-defined distinctions such as good vs. poor clinical outcome, drug response vs. non-response, or tumors with a certain histologic feature vs. tumors without that feature, may be correlated with differentially expressed genes. Unsupervised clustering, or “class discovery,” is based solely on a large data set of filtered genes, unbiased by pre-defined distinctions.

At least six analytical approaches to clustering microarray data have been described. These are hierarchical clustering (13), self-organizing maps (SOMs) (14, 15), principal component analysis (16, 17), relevance networks (18), quality threshold (QT) clustering (19, 20), and terrain maps (21). Each of these approaches posess different strengths and weaknesses in the manner in which they compare, include, and exclude data to sort out true biological signals from experimental noise.

Table 1 provides a list of links to microarray data analysis software packages that are freely available to academic researchers. Currently, the most widely used data analysis method is hierarchical clustering, a relatively fast and simple approach (13). In the first step, the two most similar elements are linked into a cluster, which is subsequently considered as a single element (or pseudoelement). This step is then repeated until the elements are sorted into a phylogenetic tree. Hierarchical clustering has the disadvantage that it employs local decision-making without considering the entire dataset (19). Moreover, the ordering of the elements of the tree is somewhat arbitrary and the tree can become extremely complex for large datasets.

Another approach to clustering microarray data is the use of SOMs, which employ an unsupervised neural-network learning algorithm (15). To generate a group of clusters, the number and geometry of clusters (or nodes) must be entered. The program then iteratively reallocates elements among these clusters until an optimum is reached. A disadvantage of this approach is that it requires input of the number of clusters (nodes) *a priori*. If too few nodes + are entered, the clusters may be highly dispersed, while if

Table 1. Links to freely available microarray data analysis software

Types of Analyses	Package Name	Link	Reference
Hierarchical clustering	Cluster/Treeview	http://genome-www5.stanford.edu/MicroArray/SMD/restech.html	13
SOM	GeneCluster	http://www-genome.wi.mit.edu/cancer/software/software.html	14
Bayesian	SAM	http://www-stat.stanford.edu/~tibs/SAM/	22
Multiple	J-Express	http://www.molmine.com/	23
Probe Modeling	dChip	http://www.dchip.org/	24
Quality Threshold	QT	http://morenoc-01.whitehead.emory.edu/qthome.html	19,20
Force-Directed	VxInsight	http://www.cs.sandia.gov/projects/VxInsight.html	21

too many nodes are entered, elements that should be grouped together can be artificially divided.

Principal component analysis attempts to replace the expression profiles of *k* genes by *p* principle components that represent large proportions of variations in gene expression patterns. The principle component is a linear combination of variables that is often visualized using a process known as multidimensional scaling to enable three-dimensional graphical representations of microarray samples (16, 25, 26). However, one problem with this approach is that it is somewhat sensitive to outliers (27).

In the fourth approach, known as relevance networks (18), pairwise comparisons are made between each member of the dataset, correlation coefficients are computed, and only those above a certain threshold are considered relevant. The threshold used by Butte *et al* (18) of 0.8 was computed by randomly permuting the data 100 times and determining that the randomly permuted data did not produce any correlations greater than 0.8. Thus, all associations with a correlation coefficient less than 0.8 were considered as possibly due to biological noise. This approach has the advantage that it can identify relationships that are negatively correlated with each other and that it considers all of the information present in the dataset.

A fifth approach used for clustering of microarray data is known as Quality Threshold (QT) clustering (19, 20). QT clustering is in many respects similar to relevance networks. In the QT clustering algorithm, each element is also compared in a pair-wise fashion and correlation coefficients are computed. Elements are clustered together such that all of the elements within a cluster must be more highly correlated with a single central element than the input quality threshold. The algorithm then computes the largest cluster that it can create using the input quality threshold, and then the next largest cluster, and so on until all of the elements are clustered together within the threshold limits.

A unique and robust clustering approach uses a force-directed energy minimization computation similar to simulated annealing algorithms. This approach minimizes distances between highly correlated genes and maximize distances between uncorrelated genes based on the t-statistic of the Pearson correlation coefficient. This algorithm, implemented in the VxInsight software developed at Sandia National Laboratories, was recently used to create "terrain maps" of global gene expression patterns in *c. elegans* (21). The robustness and utility of this clustering approach was demonstrated by showing how clusters were retained or destroyed as increasing amounts of random noise was added to the data.

A key step in any microarray data analysis is verification of statistical significance. Two of the most widely used methods for this are known as bootstrapping (28) and random permutation (18). In bootstrapping (a.k.a. cross-validation), a single gene or sample is left out of the analysis and the data is reclustered. The process is repeated for all the genes or samples, and the stability of the clusters is determined by percentage of the time that the clusters remain unchanged. In random permutation, the data labels for the samples are randomized, usually 1000 times. If a gene of a given rank in the actual data is closer to a true biological distinction than genes of the same rank in all but five of 1000 permutations, the statistical significance would be estimated at $p = 0.005$. In some studies, the number of genes that are highly correlated with random differences in samples is then compared with those highly correlated with the true biological differences in groups (e.g. clinical outcome) (29). In general, many more genes are highly correlated with true biological distinctions than with randomized groups of samples.

One statistical package developed specifically to address issues of statistical significance is the Significance Analysis of Microarrays (SAM) software (22). The SAM software uses a nonparametric empirical Bayesian model to compute a "relative difference," or d(i), which is similar to a *t*-statistic for each gene. The score assigned to each gene is based on the change in gene expression across different conditions relative to the standard deviation of repeated measurements of that gene. The genes that are called significant by SAM are those in which the observed d(i) is greater than the expected d(i) by a threshold (Δ) chosen by the user. The higher the Δ threshold, the lower the number of significant genes, and also the lower the false discovery (or false positive) rate. Genes called significant using the SAM software were shown to be much more reliably verified by Northern blot or RT-PCR, than those genes selected by a fold-change cutoff.

6. MICROARRAY INVESTIGATION OF TUMORS

In one of the first microarray studies of its kind, computer cluster analysis of gene expression in hematopoetic malignancies was able to correctly classify acute lymphoblastic and acute myeloid leukemias in all

cases studied (14). In another recent study based on analysis of 16,000 genes, 80% of 214 tumors of all different types were correctly classified into 14 different groups (30). Cluster analysis has also been used to identify subgroups of melanoma (31) lymphoma (32), breast cancer (7) and prostate cancer (6) patients. In some cases, microarray analysis has confirmed known tumor classifications, while in others cases microarray has suggested novel classification schemes. Significantly, several recent microarray studies have been predictive of outcome of breast cancers (8), prostate cancers (29), and leukemias (26).

A very recent and large study of childhood leukemias (26) examined 327 patients. 215 patients were used as a training dataset, and 112 patients were used as an independent validation dataset. Six different subtypes of pediatric acute lymphoblastic leukemias were identified, and an expression profile of 20 genes was used to classify leukemias into these subtypes with an overall accuracy of 96-100 %. In two of these subtypes, expression profiling was able to predict relapse with 97-100% accuracy.

In the breast cancer study (8), a set of 70 genes was determined to be optimal for prediction of recurrence in women with no positive lymph nodes at the time of mastectomy. An expression profile from those 70 genes correctly predicted outcome in 90% of patients in both the training set of 34 patients and in an independent set of 19 patients. Application of this classifier to lymph node negative patients could significantly reduce unnecessary adjuvant chemotherapy treatments.

Even in a smaller study of prostate cancers (29) a classifier based on five genes was more highly correlated with clinical outcome that Gleason score. The classifier reached 90% accuracy in prediction of recurrence in a set of 21 patients.

7. MICROARRAY INVESTIGATION OF BRAIN TUMORS

Cluster analysis has been applied to embryonal central nervous system tumors by Pomeroy (33). These authors investigated a total of 99 pediatric brain tumors using an Affymetrix microarray representing 6817 human genes. In general, supratentorial primitive neuroectodermal tumors (PNETs) could be distinguished from medulloblastomas, while atypical teratoid rhabdoid tumors in the brain were found to be molecularly similar to rhabdoid tumors in the abdomen. Overall, 35 out of the 42 cases studied (83%) were correctly classified. Thirty-three out of 34 medulloblastomas were correctly segregated into classic and desmoplastic subtypes compared with random classification. Among genes correlated with desmoplastic histology were PTCH, as well as three genes involved in sonic hedgehog (SHH) signaling, GLI, N-MYC and insulin-like growth factor II. These findings support the activation of SHH signaling in sporadic desmoplastic meduloblastomas, similar to the known involvement of this pathway in the desmoplastic medulloblastomas associated with Gorlin's syndrome.

Next, these authors explored the heterogeneity in clinical response to therapy (i.e., long term survival vs death) that is well known to occur in medulloblastomas. Initially, an unsupervised clustering method was used to separate 60 similarly treated medulloblastomas into two groups using self-organizing maps. Separation based on this algorithm was based predominately on differential expression of ribosome related genes, a finding confirmed by the electron microscopic demonstration of different numbers of ribosomes in the two groups. These two groups, however, showed no differences in survival. Subsequently, a supervised learning program was developed to distinguish survivors from treatment failures (i.e., death), with a minimum follow up of 24 months. Optimum predictions were made using an 8 gene model that correctly predicted outcome in 47 out of the 60 cases (78%). Several other classification algorithms performed with similar accuracy. This outcome prediction was more accurate than clinical staging and successfully separated good and poor survival groups within the TRKC negative medulloblastoma patient group (TRKC is a marker of good prognosis in medulloblastomas). According to the authors, genes associated with survival tended to be characteristic of cerebellar differentiation, while genes associated with poor outcome included genes associated with cell proliferation and metabolism, including ribosomal protein-encoding genes. Genes most frequently used by the outcome predictor as markers of survival were PLOD lysyl hydroxylase, apoD, KIAA0220, and beta NAP. Genes most commonly used as predictors of poor survival were ribosomal protein 18S, ribosomal protein L7a, MYBL2, and ribosomal protein S10.

Several groups have either limited the genes assayed to those of known function (34-36), or reported only genes of known function (10, 11, 37). Genes are generally separated into different functional groups, i.e., those involved in proliferation, invasion or progression. Not surprisingly, the majority of genes identified in these analyses have been previously studied in brain tumors. Previously studied genes that have been confirmed by microarray to be up-regulated in high-grade astrocytomas include EGFR, VEGF, transcription factor AP-2, IGF binding proteins 3 and 5, matrix metalloproteinases, TIMPs, CD44, basic fibroblast growth factor, and cathepsin H. A few genes described in systemic cancers but not previously related to brain tumors have been identified, including insulin growth factor binding protein 2 and apolipoprotein D. Other genes important in the oncogenesis of brain tumors may remain hidden in databases, their significance unrecognized by current data interpretation techniques.

In one of the initial microarray studies of brain tumors, Fuller analysed various grades of fibrillary astrocytomas using a Clontech microarray plate containing 588 cancer related genes(35). Insulin growth factor binding protein 2 (IGFBP2) was consistently over expressed only in the glioblastomas, i.e., the highest grade of infiltrating astrocytomas, suggesting that it is a marker of progression in fibrillary astrocytomas. Sallienen, also using a 588 gene Clontech microarray, found 117 up

regulated genes in glioblastomas relative to normal brain, the most highly over expressed gene being IGFBP 2(36). Anaplatic (WHO grade III) and low-grade (WHO grade II) infiltrating astrocytomas over expressed 32 and 38 genes respectively. These authors also used a tissue array of 418 astrocytomas to confirm the association of IGFBP2 with astrocytoma progression at the protein level.

IGFBPs are a family of proteins that bind insulin growth factors with high affinity. These factors bind to specific cell surface receptors as well as the extracellular matrix and exert both insulin growth factor-dependent and -independent functions. IGFBPs may either suppress or augment insulin growth factor activity, depending upon experimental conditions. IGFBP2 has been associated with several different malignant tumors, including prostate and ovarian cancer (38, 39). Preliminary studies using cell lines transfected with IGFBP2 expressing vectors suggest that this protein may play a role in either tumor growth or invasion (40-42). This protein is expressed in highly proliferative fetal tissues, including astrocytes, with significantly decreased expression after birth (43).

Huang studied 1176 cancer associated genes in eleven low-grade infiltrating astrocytomas (WHO grade II)(34). Thirteen genes were found to be up regulated in 20 to 100% of cases, while eleven genes were down regulated. Eleven genes were further tested by RT-PCR, nine of which were confirmed to be differentially expressed. Two of the identified genes were not differentially expressed by RT-PCR, presumably representing false positives. SPARC, previously associated with infiltrating astrocytomas, was up-regulated in 30% of cases by microarray; however, the SPARC protein was increased by immunohistochemistry in 100% of cases. This finding illustrates the general point that mRNA and protein levels do not always correlate directly. Tissue inhibitor of matrix metalloproteinase (TIMP3) was over expressed in all low-grade astrocytomas. TIMP3 is involved with modulation of the extracellular matrix and may also play a role in the invasive behavior of astrocytic neoplasms(44). No differences in expression were detected between the eight low-grade astrocytomas with p53 mutations and the three low-grade astrocytomas without p53 mutations.

SPARC, a secreted glycoprotein, resides in the extracellual matrix (45). SPARC is highly expressed in a variety of neoplasms, including, colon cancer, breast cancer, ovarian cancer, melanoma, and meningioma, as well as in non-neoplastic tissue repair. Similar to IGFBP2, it is highly expressed in fetal tissue, including astrocytes and blood vessels (46). In addition, SPARC has been previously shown to be present in all grades of fibrillary astrocytomas. It has been hypothesized that SPARC plays a role in the invasion of astrocytic neoplasms (47).

Rickman, using the Affymetrix system, analysed the expression of 6800 genes in glioblastomas (grade IV infiltrating astrocytomas) relative to pilocytic and grade 2 infiltrating astrocytomas (11). Three hundred-sixty genes were differentially expressed in glioblastomas relative to pilocytic astrocytomas. Five genes not previously assocated with glioblastomas, syndecan-1, filamin A, fork head box M1, fork head box G1b and ZYX, were chosen for additional study. Differential expression of all five genes was confirmed by quantitative PCR. In this study, expression microarray tended to underestimate the level of differential expression, an observation that has also been made by other investigators. IGFBP 2 was up-regulated in glioblastomas, while apoD was up-regulated in the pilocytic tumors.

Ljubimova analysed different grades of fibrillary astrocytomas using a microarray from Incyte containing 11,000 genes (37). In contrast with the Rickman's results, only 14 genes were consistently over expressed in glioblastomas relative to normal brain tissue. It would appear that not all microarrays have the same sensitivity. Most of the 14 genes identified were previously known to be expressed in glioblastomas, with the exception of the alpha4 laminin chain. Alpha4 laminin chain protein was identified in blood vessel walls by immunohistochemistry, suggesting that this gene is associated with the angiogenic response in glioblastomas.

Comparing the differential expression of 6800 genes in pilocytic and anaplastic astrocytomas, Hunter identified apolipoprotein D as the most highly over expressed marker of pilocytic tumors (10). Subsequent immunohistochemical studies showed that apoD immunostaining was associated not only with pilocytic astrocytomas, but also with other low-grade and potentially curable primary brain tumors, including gangliogliomas, subependymal giant cell astrocytomas, and a single pleomorphic xantoastrocytoma. In contrast, all of the fatal fibrillary astrocytomas were negative for apoD immunostaining. Positive apoD immunostainining in the low-grade glial tumors was present in cytoplasm and cyst fluid. Positive immunostaining was also present in granular bodies, a characteristic morphologic feature of pilocytic astrocytomas, gangliogliomas, and pleomorphic xantoastrocytomas. This study, together with the identification of apoD as a prognostic marker in medulloblastomas by Pomeroy (33), highlight apoD as a promising marker of prognosis in primary brain tumors.

ApoD is a member of the lipocalin family of proteins involved in the transport of small hydrophobic molecules (48). ApoD is a component of high density lipoproteins in serum and is the protein with the highest concentration in fluid from benign cysts of the breast (49). Similar to IGFBP 2, apoD has been previously identified as a prognostic marker in several types of carcinoma outside of the brain (50, 51). In breast and prostate carcinoma cell lines, as well as senescent fibroblasts, apoD has been proposed as a marker of cell cycle arrest (52-54).

Gutmann *et al.* used the Affymetrix system to study the expression of 11,000 genes in 8 pilocytic astrocytomas, 3 oligodendrogliomas, 3 normal white matter specimens and 2 NHA specimens (normal human astrocytes from fetuses)(9). Hierarchical cluster analysis clearly delineated the pilocytic astrocytomas from both oligodendrogliomas and normal white matter. The overall

expression pattern of the pilocytic tumors most closely resembled that of NHA cells; however, pilocytic tumors also expressed markers characteristic of oligodendrocytes, such as myelin basic protein, PMP-22, and proteolipid protein. Western blots confirmed over expression of N-CAM (neural cell adhesion molecule) in pilocytic tumors compared with NHA cells and under expression of connexin-43 in pilocytic tumors relative to both NHA cells and oligodendrogliomas. Both apoD and SPARC were up-regulated in pilocytic astrocytomas relative to NHA cells. Although unconfirmed, one potential gene, EF-1alpha2, was identified which might distinguish NF-1 associated from sporadic pilocytic astrocytomas.

8. CONCLUSION

Gene expression is fundamental to the determination of cell behavior and the process of malignant transformation. Each malignant cell type expresses an estimated 10,000 mRNA species, and a subset of these mRNA species send a still incompletely understood message signaling malignant behavior. The capacity of expression microarrays to simultaneously assay all mRNA species expressed by the human genome has understandably sparked great excitement in the field of oncology. However, analysis of the massive amounts of data generated by microarrays is daunting. Separating true biologic signal from technical noise may require large numbers of cases and sophisticated statistical analysis. Because of post-translational modifications and complex systems of protein activation, mRNA levels often correlate poorly with protein levels or activity. Translating the malignant message encoded in the mRNA of malignant cells represents a challenge for the future.

In the few years of its existence, microarray technology has already contributed to the study of human brain tumors. Prognostic groups within medulloblastoma patients have been separated. Medulloblastomas and supratentorial PNETs, morphologically similar neoplasms, have been distinguished from each other. The roles of numerous known cancer related genes in brain tumors have been confirmed. Insulin growth factor 2 and apolipoprotein D have been identified as novel players in primary brain neoplasia. Even better results will undoubtedly come with technical improvements, more appropriate reference mRNA sources, better data analysis methods, larger data sets, and increased knowledge concerning the functions and interactions of genes.

9. REFERENCES

1. Hanahan D, RA Weinberg: The hallmarks of cancer. *Cell* 100 (1), 57-70 (2000)
2. Brazma A, P Hingamp, J Quackenbush, G Sherlock, P Spellman, C Stoeckert, J Aach, W Ansorge, CA Ball, HC Causton, T Gaasterland, P Glenisson, FC Holstege, IF Kim, V Markowitz, JC Matese, H Parkinson, A Robinson, U Sarkans, S Schulze-Kremer, J Stewart, R Taylor, J Vilo, M Vingron: Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 29 (4), 365-71 (2001)
3. Ohyama H, X Zhang, Y Kohno, I Alevizos, M Posner, DT Wong, R Todd: Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. *Biotechniques* 29 (3), 530-6 (2000)
4. Wang E, LD Miller, GA Ohnmacht, ET Liu, FM Marincola: High-fidelity mRNA amplification for gene profiling. *Nat Biotechnol* 18 (4), 457-9 (2000)
5. Luo L, RC Salunga, H Guo, A Bittner, KC Joy, JE Galindo, H Xiao, KE Rogers, JS Wan, MR Jackson, MG Erlander: Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* 5 (1), 117-22 (1999)
6. Dhanasekaran SM, TR Barrette, D Ghosh, R Shah, S Varambally, K Kurachi, KJ Pienta, MA Rubin, AM Chinnaiyan: Delineation of prognostic biomarkers in prostate cancer. *Nature* 412 (6849), 822-6 (2001)
7. Perou CM, T Sorlie, MB Eisen, M van de Rijn, SS Jeffrey, CA Rees, JR Pollack, DT Ross, H Johnsen, LA Akslen, O Fluge, A Pergamenschikov, C Williams, SX Zhu, PE Lonning, AL Borresen-Dale, PO Brown, D Botstein: Molecular portraits of human breast tumours. *Nature* 406 (6797), 747-52. (2000)
8. van 't Veer LJ, H Dai, MJ van de Vijver, YD He, AA Hart, M Mao, HL Peterse, K van der Kooy, MJ Marton, AT Witteveen, GJ Schreiber, RM Kerkhoven, C Roberts, PS Linsley, R Bernards, SH Friend: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415 (6871), 530-6. (2002)
9. Gutmann DH, NM Hedrick, J Li, R Nagarajan, A Perry, MA Watson: Comparative gene expression profile analysis of neurofibromatosis 1- associated and sporadic pilocytic astrocytomas. *Cancer Res* 62 (7), 2085-91. (2002)
10. Hunter S, A Young, J Olson, DJ Brat, G Bowers, JN Wilcox, D Jaye, S Mendrinos, A Neish: Differential expression between pilocytic and anaplastic astrocytomas: identification of apolipoprotein D as a marker for low-grade, non- infiltrating primary CNS neoplasms. *J Neuropathol Exp Neurol* 61 (3), 275-81. (2002)
11. Rickman DS, MP Bobek, DE Misek, R Kuick, M Blaivas, DM Kurnit, J Taylor, SM Hanash: Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Res* 61 (18), 6885-91 (2001)
12. Novak JP, R Sladek, TJ Hudson: Characterization of variability in large-scale gene expression data: implications for study design. *Genomics* 79 (1), 104-13 (2002)
13. Eisen MB, PT Spellman, PO Brown, D Botstein: Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95 (25), 14863-8 (1998)
14. Golub TR, DK Slonim, P Tamayo, C Huard, M Gaasenbeek, JP Mesirov, H Coller, ML Loh, JR Downing, MA Caligiuri, CD Bloomfield, ES Lander: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286 (5439), 531-7 (1999)
15. Tamayo P, D Slonim, J Mesirov, Q Zhu, S Kitareewan, E Dmitrovsky, ES Lander, TR Golub: Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci U S A* 96 (6), 2907-12 (1999)
16. Lu J, Z Liu, M Xiong, Q Wang, X Wang, G Yang, L Zhao, Z Qiu, C Zhou, M Wu: Gene expression profile changes in initiation and progression of squamous cell

carcinoma of esophagus. *Int J Cancer* 91 (3), 288-94 (2001)
17. Hilsenbeck SG, WE Friedrichs, R Schiff, P O'Connell, RK Hansen, CK Osborne, SA Fuqua: Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J Natl Cancer Inst* 91 (5), 453-9 (1999)
18. Butte AJ, P Tamayo, D Slonim, TR Golub, IS Kohane: Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci U S A* 97 (22), 12182-6 (2000)
19. Heyer LJ, S Kruglyak, S Yooseph: Exploring expression data: identification and analysis of coexpressed genes. *Genome Res* 9 (11), 1106-15 (1999)
20. Young AN, MB Amin, CS Moreno, SD Lim, C Cohen, JA Petros, FF Marshall, AS Neish: Expression profiling of renal epithelial neoplasms: a method for tumor classification and discovery of diagnostic molecular markers. *Am J Pathol* 158 (5), 1639-51 (2001)
21. Kim SK, J Lund, M Kiraly, K Duke, M Jiang, JM Stuart, A Eizinger, BN Wylie, GS Davidson: A gene expression map for Caenorhabditis elegans. *Science* 293 (5537), 2087-92 (2001)
22. Tusher VG, R Tibshirani, G Chu: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 98 (9), 5116-21 (2001)
23. Dysvik B, I Jonassen: J-Express: exploring gene expression data using Java. *Bioinformatics* 17 (4), 369-70 (2001)
24. Li C, WH Wong: Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A* 98 (1), 31-6 (2001)
25. Luo J, DJ Duggan, Y Chen, J Sauvageot, CM Ewing, ML Bittner, JM Trent, WB Isaacs: Human prostate cancer and benign prostatic hyperplasia: molecular dissection by gene expression profiling. *Cancer Res* 61 (12), 4683-8 (2001)
26. Yeoh E-J, ME Ross, SA Shurtleff, WK Williams, D Patel, R Mahfouz, FG Behm, SC Raimondi, MV Relling, A Patel, C Cheng, D Campana, D Wilkins, X Zhou, J Li, H Liu, C-H Pui, WE Evans, C Naeve, L Wong, JR Downing: Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1 (2), 133-143 (2002)
27. Wittes J, HP Friedman: Searching for evidence of altered gene expression: a comment on statistical analysis of microarray data. J Natl Cancer Inst 91 (5), 400-1 (1999)
28. Kerr MK, GA Churchill: Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc Natl Acad Sci U S A* 98 (16), 8961-5 (2001)
29. Singh D, PG Febbo, K Ross, DG Jackson, J Manola, C Ladd, P Tamayo, AA Renshaw, AV D'Amico, JP Richie, ES Lander, M Loda, PW Kantoff, TR Golub, WR Sellers: Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* 1 (2), 203-209 (2002)
30. Ramaswamy S, P Tamayo, R Rifkin, S Mukherjee, CH Yeang, M Angelo, C Ladd, M Reich, E Latulippe, JP Mesirov, T Poggio, W Gerald, M Loda, ES Lander, TR Golub: Multiclass cancer diagnosis using tumor gene expression signatures. *Proc Natl Acad Sci U S A* 98 (26), 15149-54 (2001)
31. Bittner M, P Meltzer, Y Chen, Y Jiang, E Seftor, M Hendrix, M Radmacher, R Simon, Z Yakhini, A Ben-Dor, N Sampas, E Dougherty, E Wang, F Marincola, C Gooden, J Lueders, A Glatfelter, P Pollock, J Carpten, E Gillanders, D Leja, K Dietrich, C Beaudry, M Berens, D Alberts, V Sondak: Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* 406 (6795), 536-40 (2000)
32. Alizadeh AA, MB Eisen, RE Davis, C Ma, IS Lossos, A Rosenwald, JC Boldrick, H Sabet, T Tran, X Yu, JI Powell, L Yang, GE Marti, T Moore, J Hudson, Jr., L Lu, DB Lewis, R Tibshirani, G Sherlock, WC Chan, TC Greiner, DD Weisenburger, JO Armitage, R Warnke, LM Staudt, et al.: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403 (6769), 503-11 (2000)
33. Pomeroy SL, P Tamayo, M Gaasenbeek, LM Sturla, M Angelo, ME McLaughlin, JY Kim, LC Goumnerova, PM Black, C Lau, JC Allen, D Zagzag, JM Olson, T Curran, C Wetmore, JA Biegel, T Poggio, S Mukherjee, R Rifkin, A Califano, G Stolovitzky, DN Louis, JP Mesirov, ES Lander, TR Golub: Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 415 (6870), 436-42 (2002)
34. Huang H, S Colella, M Kurrer, Y Yonekawa, P Kleihues, H Ohgaki: Gene expression profiling of low-grade diffuse astrocytomas by cDNA arrays. *Cancer Res* 60 (24), 6868-74 (2000)
35. Fuller GN, CH Rhee, KR Hess, LS Caskey, R Wang, JM Bruner, WK Yung, W Zhang: Reactivation of insulin-like growth factor binding protein 2 expression in glioblastoma multiforme: a revelation by parallel gene expression profiling. *Cancer Res* 59 (17), 4228-32 (1999)
36. Sallinen SL, PK Sallinen, HK Haapasalo, HJ Helin, PT Helen, P Schraml, OP Kallioniemi, J Kononen: Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. *Cancer Res* 60 (23), 6617-22 (2000)
37. Ljubimova JY, AJ Lakhter, A Loksh, WH Yong, MS Riedinger, JH Miner, LM Sorokin, AV Ljubimov, KL Black: Overexpression of alpha4 chain-containing laminins in human glial tumors identified by gene microarray analysis. *Cancer Res* 61 (14), 5601-10 (2001)
38. Ho PJ, RC Baxter: Insulin-like growth factor-binding protein-2 in patients with prostate carcinoma and benign prostatic hyperplasia. *Clin Endocrinol* (Oxf) 46 (3), 333-42 (1997)
39. Karasik A, J Menczer, C Pariente, H Kanety: Insulin-like growth factor-I (IGF-I) and IGF-binding protein-2 are increased in cyst fluids of epithelial ovarian cancer. *J Clin Endocrinol Metab* 78 (2), 271-6 (1994)
40. Zhang W, H Wang, SW Song, GN Fuller: Insulin-like growth factor binding protein 2: gene expression microarrays and the hypothesis-generation paradigm. *Brain Pathol* 12 (1), 87-94 (2002)
41. Hoeflich A, O Fettscher, H Lahm, WF Blum, HJ Kolb, D Engelhardt, E Wolf, MM Weber: Overexpression of insulin-like growth factor-binding protein-2 results in increased tumorigenic potential in Y-1 adrenocortical tumor cells. *Cancer Res* 60 (4), 834-8 (2000)

42. Menouny M, M Binoux, S Babajko: IGFBP-2 expression in a human cell line is associated with increased IGFBP-3 proteolysis, decreased IGFBP-1 expression and increased tumorigenicity. *Int J Cancer* 77 (6), 874-9 (1998)
43. Lee WH, KM Michels, CA Bondy: Localization of insulin-like growth factor binding protein-2 messenger RNA during postnatal brain development: correlation with insulin-like growth factors I and II. *Neuroscience* 53 (1), 251-65 (1993)
44. Uhm JH, NP Dooley, JG Villemure, VW Yong: Glioma invasion in vitro: regulation by matrix metalloprotease-2 and protein kinase C. *Clin Exp Metastasis* 14 (5), 421-33 (1996)
45. Swaroop A, BL Hogan, U Francke: Molecular analysis of the cDNA for human SPARC/osteonectin/BM-40: sequence, expression, and localization of the gene to chromosome 5q31- q33. *Genomics* 2 (1), 37-47 (1988)
46. Mundlos S, B Schwahn, T Reichert, B Zabel: Distribution of osteonectin mRNA and protein during human embryonic and fetal development. *J Histochem Cytochem* 40 (2), 283-91 (1992)
47. Rempel SA, WA Golembieski, S Ge, N Lemke, K Elisevich, T Mikkelsen, JA Gutierrez: SPARC: a signal of astrocytic neoplastic transformation and reactive response in human primary and xenograft gliomas. *J Neuropathol Exp Neurol* 57 (12), 1112-21 (1998)
48. Flower DR: The lipocalin protein family: a role in cell regulation. *FEBS Lett* 354 (1), 7-11. (1994)
49. Balbin M, JM Freije, A Fueyo, LM Sanchez, C Lopez-Otin: Apolipoprotein D is the major protein component in cyst fluid from women with human breast gross cystic disease. *Biochem J* 271 (3), 803-7 (1990)
50. Diez-Itza I, F Vizoso, AM Merino, LM Sanchez, J Tolivia, J Fernandez, A Ruibal, C Lopez-Otin: Expression and prognostic significance of apolipoprotein D in breast cancer. *Am J Pathol* 144 (2), 310-20 (1994)
51. Vazquez J, L Gonzalez, A Merino, F Vizoso: Expression and clinical significance of apolipoprotein D in epithelial ovarian carcinomas. *Gynecol Oncol* 76 (3), 340-7 (2000)
52. Provost PR, YL Marcel, RW Milne, PK Weech, E Rassart: Apolipoprotein D transcription occurs specifically in nonproliferating quiescent and senescent fibroblast cultures. *FEBS Lett* 290 (1-2), 139-41 (1991)
53. Lopez-Boado YS, J Tolivia, C Lopez-Otin: Apolipoprotein D gene induction by retinoic acid is concomitant with growth arrest and cell differentiation in human breast cancer cells. *J Biol Chem* 269 (43), 26871-8 (1994)
54. Simard J, R Veilleux, Y de Launoit, DE Haagensen, F Labrie: Stimulation of apolipoprotein D secretion by steroids coincides with inhibition of cell proliferation in human LNCaP prostate cancer cells. *Cancer Res* 51 (16), 4336-41 (1991)

Key Words: Brain, Tumors, Microarray, Astrocytomas, mRNA, Review

Send correspondence to: Dr Stephen Hunter, Emory University School of Medicine, Department of Pathology; Room H-173, Atlanta Georgia 30087, Tel: 404-712-4278, Fax: 404-712-4754, E-mail: Stephen_Hunter@Emory.org