Preface

Advances in our knowledge of the molecular basis of cancer are at the heart of the present revolution in clinical oncology. The identification of tumor-specific molecular alterations has led to new means of diagnosis and classification, and the characterization of critical pathways regulating tumor growth is providing the potential for less toxic, more effective targeted therapy. Nonetheless, these advances had previously occurred at an agonizingly slow pace, i.e., one gene at a time. That investigative pace has now been dramatically altered by the completion of a draft of the entire human genome and the development of miniaturized high-throughput technology for genetic analysis. These extraordinary accomplishments now permit not only the monitoring of every gene sequence in a single experiment, but also a comprehensive analysis of the complex coordinated programs and pathways that contribute to the clinical phenotype of cancers. This rapid and comprehensive approach to the investigation of tumor biology has the potential to dramatically shape the future of clinical oncology.

Expression Profiling of Human Tumors: Diagnostic and Research Applications is intended to provide an introduction and overview to comprehensive gene expression profiling of human tumors, one of the most promising new high-throughput investigative approaches in molecular biology. The intent was to provide not only a primer for the technology and analytical methods, but also an early assessment of the state-of-the-art with respect to both successes and pitfalls. These successes are significant and include methods of more precise diagnosis, and identification of prognostic markers, therapeutic targets, and gene expression patterns that predict therapeutic response. Nonetheless, there are significant challenges to further success, such as procurement and processing of appropriate samples, improvement and validation of technical approaches, and refinement of analytical methods for the resulting complex datasets. We have attempted to provide a balance between the basic science aspects of this work and its application to the clinical setting, but we have focused on the analysis of human tissue samples as providing the most direct means of translating findings to clinical practice. There are many complex issues that need to be considered as this type of work goes forward, and we hope this text will serve as a starting point for future discoveries.

The emphasis here on gene expression profiling is not intended to suggest that this should be considered the ultimate view of the molecular biology of the cancer cell. On the contrary, we all look forward to the day when analysis at the protein level is as comprehensive and provides as much detail as the present attempts of global gene transcript measurements. Obviously, the closer we come to assessment of the actual function of each molecule, the more accurate our abilities to correlate those with the clinical phenotype. Proteomics holds the promise to better achieve that goal, but is still in its

infancy, with even greater hurdles to overcome than we presently face with sequence-based expression analysis. We leave that topic for future publications.

We would like to express our deep appreciation to the many authors who have provided overviews of work in their fields. These individuals have contributed their time and effort to provide highly useful information for others (sometimes while being badgered by the Editors!). We would also like to thank Ms. Fabienne Volel and Ms. Shirley Tung for excellent assistance. Finally we thank our families for their patience and support.

Expression Profiling of Human Tumors: Diagnostic and Research Applications clearly depicts the rapid advances that are occurring in clinically important areas and that will no doubt increasingly impact clinical care. We sincerely hope that our book provides information useful to all basic or clinical investigators concerned with the molecular basis of cancer and the improvement of cancer care.

Marc Ladanyi, ***MD***
William L. Gerald, ***MD, PhD***

2
cDNA Microarrays

Paul S. Meltzer

INTRODUCTION

Cancer can be viewed as a disease of disturbed genome function. The phenomena of aberrant growth, differentiation, invasion, and metastasis are the phenotypic manifestations of an underlying genetic process. Ultimately, irrespective of whether this is the result of point mutation, translocation, deletion, gene amplification, or methylation, the malignant phenotype is mediated by a characteristic pattern of gene expression. Identifying the genes whose expression differs between normal tissues and tumors and among tumor types has long been a focus of cancer researchers. This endeavor was tremendously accelerated by the development of technologies for the parallel analysis of gene expression. This chapter will focus on one of these technologies, cDNA microarrays. The ability to measure the expression of tens of thousands of genes in a tumor specimen has revolutionized our ability to describe cancers. A rapidly burgeoning literature offers hope that this improvement will translate into improved diagnosis and prognosis, as well as accelerate the discovery of new therapeutic targets.

The pivotal concept enabling cDNA microarray technology is simple. Rather than maintaining libraries of cDNA clones as stocks of bacteria mixed in suspension, libraries can be stored as collections of individual clones arrayed in microtiter plates. This essential aspect of expressed sequence tag (EST) library sequencing projects provides a residual physical resource that can be used for other purposes *(1)*. Libraries in this format can be screened for individual genes of interest by replacing the traditional colony lift with a filter prepared by transferring bacteria from the source plates to a hybridization membrane *(2)*. Such filters can also be hybridized with labeled cDNA prepared from a cell source of interest *(3,4)*. By quantitating the hybridization signal, an estimate of the expression of the gene corresponding to each cDNA can be obtained. The cDNA microarray, which has now found wide use in all fields of biomedical research, is the much refined descendant of this simple concept. The fundamental elements, which are necessary to carry out this analysis, are arrayed libraries of cDNA clones, a means for producing hybridizable arrays of these cDNAs, a system to detect hybridization signal, and a means to quantitate those signals link them to the individual cDNAs and compare these data across sample sets. The following discussion will briefly consider the individual elements of the system and the considerations in experimental design that are of particular relevance to cancer research.

From: *Expression Profiling of Human Tumors: Diagnostic and Research Applications*
Edited by: Marc Ladanyi and William L. Gerald

cDNA LIBRARIES

One of the great attractions of the cDNA microarray platform is its flexibility, and in principle, one can construct arrays from any cDNA library that the investigator might select. One could construct clone arrays that represent specific pathways or protein classes or even use nonsequenced libraries from a tissue of interest. In practice, most researchers use clones that have been culled from EST sequencing projects. There are now over 4.4 million sequences in the National Center for Biotechnology Information (NCBI) database of ESTs. Given that the number of genes in the human genome is two orders of magnitude lower, it is apparent that there is considerable redundancy in the EST sequence data. Using the EST sequence data and the mRNA sequences of known genes, bioinformatic tools have been used to cluster sequences into groups representing individual transcripts. The most widely used system, Unigene, is maintained by NCBI *(5,6)*. Each individual cluster is designated by an identifier that can be used to extract the set of sequences that constitute that cluster. Clone sets are selected to represent each Unigene cluster. These clones must then be physically retrieved from their source plates and rearrayed into sets for microarray fabrication. Ideally, each clone is sequence-verified at the time it is rearrayed to maintain a high standard for sequence authenticity in the final rearrayed library.

Each strategy for microarray production has intrinsic strengths and weaknesses. Ultimately, one would like to have the option of constructing microarrays that include a complete representation of the genome. To accomplish this, it would be necessary to retrieve a cDNA clone for each gene, a goal that is limited by a number of factors, including the still incomplete annotation of the human genome sequence. It is relatively easy to access the genes that have been encountered multiple times in the course of sequencing EST libraries. However, although over 800 libraries have contributed data to the Unigene database, some genes, which may be expressed only in specialized tissues or at developmental stages that have not been sampled, may not be represented at all. Genes that have been sequenced only a few times may be difficult to locate, depending on how effectively libraries have been archived. These considerations have not posed major limitations for expression profiling studies of human cancers, but the possibility that key genes may be missing from a given microarray is important to bear in mind when considering the results of any study. In addition, the Unigene clustering system undergoes periodic revision (builds) as new data becomes available, so clusters are not stable over time. There are also significantly more clusters (over 100,000) than the estimated number of human genes, and there is certainly both noise (due to artifactual cDNA clones) and redundancy (multiple clusters for the same gene) within Unigene. Over 36,000 clusters are represented by only a single sequence. These are difficult to access and may include clones that represent library artifacts or genes with very low expression.

Another limitation inherent to cDNA libraries is the problem of preserving sequence authenticity. In general, for microarray applications, libraries of rearrayed sequence-verified clones are used. However, in the manipulation of tens of thousands of bacterial stocks, it is inevitable that a residual level of error remains, and investigators must bear this caveat in mind. Despite all these difficulties, cDNAs have major attractions. They are readily available at a relatively low cost and can be manipulated with familiar tech-

niques. Once clones are obtained, an unlimited supply of DNA for printing can be obtained by polymerase chain reaction (PCR), and the clones themselves are a convenient source of probes for follow-up studies. The cDNA technology lends itself to specialized projects potentially utilizing special purpose libraries constructed from material of interest to an investigator and potentially enriched for disease-specific genes, which might not be included in generic clone collections. Finally, of the various expression microarray technologies, only cDNA arrays lend themselves to the determination of gene copy number by comparative genomic hybridization, an analysis that adds a potentially important dimension to tumor profiling studies *(7,8)*.

How big does a cDNA microarray have to be to generate useful information for tumor profiling? It is quite clear that full genome-scale arrays are not necessary, as the world literature to date falls short of this level. Most investigators conclude that they would like to use the largest available array, because the analysis is a destructive process, and sample sets may be more limiting than the arrays themselves. However, although this issue has not been studied systematically, there seems to be a decline in useful information as genes are added to arrays. If one imagines a list of genes ranked as to their frequency of expression in tissues, the lower portion of this list will contain genes that are very infrequently expressed and, therefore, are less likely to be expressed in any given tissue of interest. This tends to counterbalance the tendency of cDNA clone sets to be limited to the 10,000–20,000 Unigene clusters representing the most commonly expressed genes.

CONSTRUCTING MICROARRAYS

Once a clone set has been selected, fabricating microarrays is quite straightforward. The technology is dependent on the use of a robotic device to deposit DNA (typically a PCR product) from each clone on a solid support, usually a glass microscope slide *(9)*. As an alternative to glass, microarrays can also be printed on nylon membranes for use in radioactive detection systems rather than the fluorescence-based detection used for glass microarrays. Detailed protocols for cDNA spotting are readily available. Robots for printing microarrays are produced by several manufacturers. The printing procedure is sufficiently simple that many institutions have established facilities for constructing microarrays, and expertise in array fabrication is now quite widespread. Commercial sources of spotted cDNAs are now available and represent an alternative to locally fabricated microarrays. It should be noted that once a spotting facility has been established, only minor modifications are necessary to spot alternative DNAs, such as synthetic oligonucleotides.

There is one important consideration that investigators who plan to use microarrays for tumor classification should bear in mind. It is extremely important that data sets, which are designed to provide this type of information, be generated in as homogeneous a fashion as possible. This maximizes the possibility of recognizing smaller differences in expression between sample subgroups and minimizes the number of false positives due to nonuniformities in technique. Of the many sources of this type of error, slide-to-slide variation is perhaps the most important. Generally, within a print batch, this error is relatively small and well compensated for by the use of a two-color hybridization scheme. However, when comparing batches of slides printed at different times, a large number of variables can interact to result in significant nonuniformities

between batches. This can present a problem, which is relevant to project design. For example, if a given printing system can generate a batch of 100 slides, then no more than 100 specimens can be compared within a single batch. Switching to a second batch of slides for the next 100 specimens may yield data that can still be useful, but it will not be as satisfactory for recognizing the smaller differences between samples, which may be of the greatest interest to clinical investigators. This difficulty can potentially be overcome by improvements in printing technique, but these are most likely to take place within an industrial environment. This issue, at the very least, deserves consideration in the design of tumor profiling studies using cDNA microarrays.

MEASURING GENE EXPRESSION ON cDNA MICROARRAYS (FIG. 1)

In order to generate the primary expression data, a labeled representation of the sample mRNAs must be prepared for hybridization to the microarray. Each feature on the array is referred to as a "probe," and the mixture derived from the sample is the "target." Fluorescence detection has emerged as the most useful methodology when coupled with the use of glass microarrays *(9)*. Fluorescence allows for simultaneous hybridization of an unknown and a reference sample, each labeled with distinct fluorochromes. This forgives, to a large extent, any imperfections in array fabrication and allows a very accurate and sensitive measurement of the unknown relative to the reference source. As an alternative, radiolabeled targets can be hybridized to nylon membrane microarrays. This presents some difficulties in image analysis, but is a viable alternative if access to glass arrays is not possible.

For hybridization, the mRNA from the sample is converted to a labeled derivative by reverse transcription to cDNA. A modified nucleotide is included in the cDNA synthesis reaction. A fluorochrome can be incorporated directly, coupled to a reactive group (as in the aminoallyl labeling strategy), or used in secondary detection. The dynamic range and signal intensity are two of the critical variables affecting labeling methods. Investigators using tissue samples prefer to minimize sample requirements. The direct incorporation of a fluorescent dye requires 20–100 μg of total RNA, while aminoallyl labeling requires 1–20 μg RNA. These are quantities that are easy to achieve with small tissue specimens. The use of smaller samples requires an amplification step. This can be accomplished by incorporating one or more cycles of in vitro transcription driven from a bacteriophage RNA polymerase promoter incorporated in the primer used for cDNA synthesis. Using this approach, useful data has been obtained from minute quantities of RNA *(10,11)*. Investigators using amplification techniques should be aware that consistent labeling techniques should be used for a given project. Microdissection, with comparison of tumor and normal cells, is particularly attractive as an approach to directly identify genes that are cancer- rather than tissue-specific in their expression pattern *(12,13)*.

For two-color hybridization, it is necessary to select a reference sample. In principle, the primary requirement of this material is a similar pattern of gene expression to the tumors for which it will be compared. If many genes, which are strongly expressed in the tumors, are expressed in the reference sample at near background levels, then the sample-to-reference ratio will be unreliable. This requirement for similar expression may be difficult to meet. One approach is to use a related cancer cell line or, as an alternative, a pool of cell lines. There are distinct advantages to using a pool. Specifi-

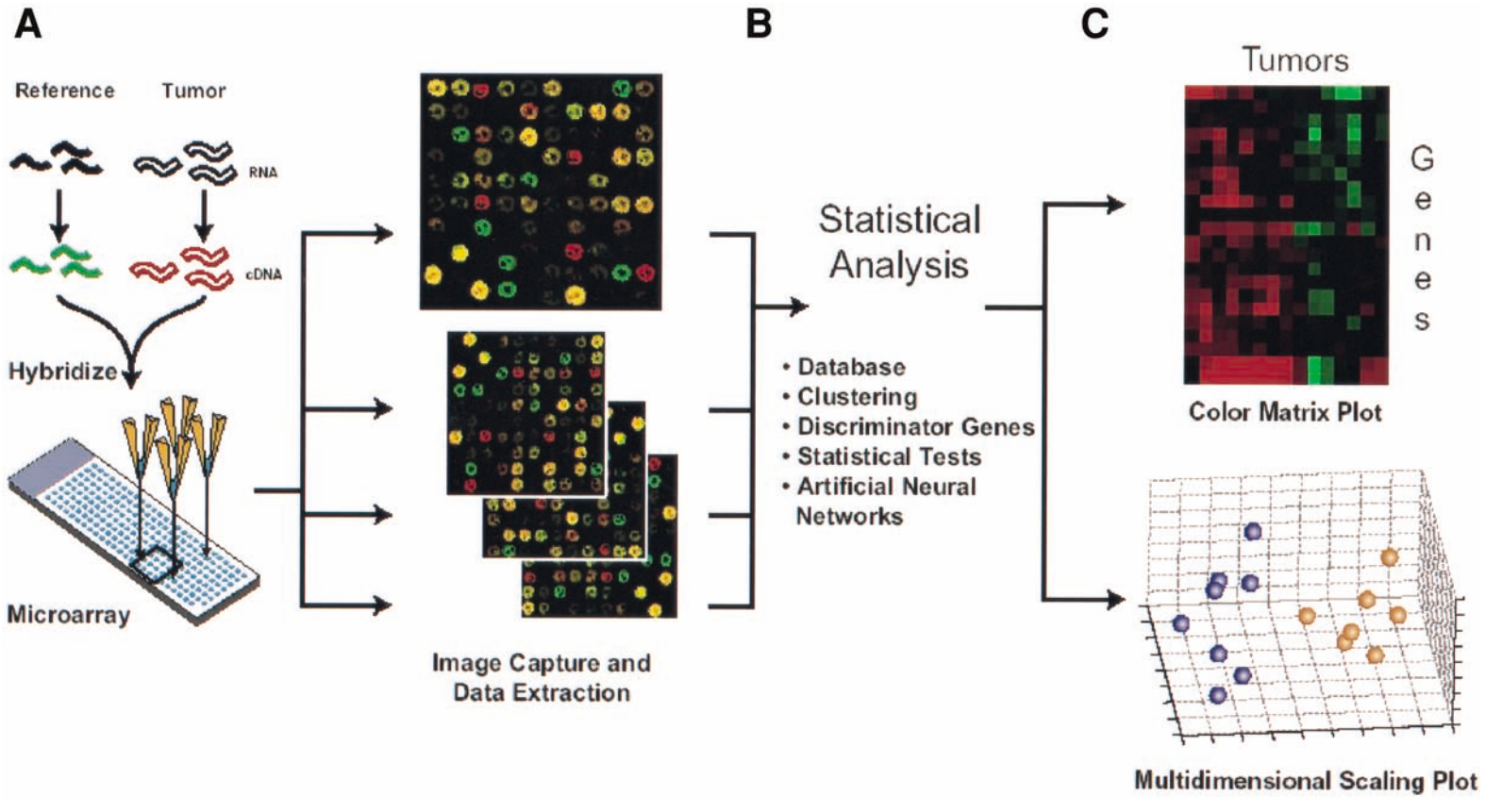

Fig. 1. cDNA microarray data flow. **(A)** Representations of the cellular mRNA pool are prepared by reverse transcription and hybridized to a robotically printed microarray. Each spot on the array represents a cDNA clone that can be assigned to a specific gene. Microarrays are typically hybridized with a mixture of tumor and reference samples, each labeled with a distinct fluorochrome. **(B)** After hybridization, the fluorescent images are captured in a scanner, and the quantitative hybridization signals are extracted for each array element using image analysis software. **(C)** Data from a series of samples is stored in a relational database, analyzed with statistical methods appropriate to the research question posed in the study, and displayed for inspection. Two of several available types of data display are used.

cally, each component of the pool will eliminate some low denominator genes, and variations between batches of cells are minimized across the pool. Ideally, a project is not started until sufficient reference RNA is available to complete the entire project. It is important to carry out test hybridizations to determine the suitability of a reference RNA before proceeding. Normal tissue truly representing the cancer progenitor cell is not generally available in sufficient quantity for use as a reference. The exception may be those situations where microdissected material will be amplified and where flanking normal cells might also be obtained for similar processing.

A recurring question in tumor processing is the influence of admixed normal stromal, endothelial, and inflammatory cells on the pattern of gene expression. Because it is significantly easier to generate data from whole tissues compared to microdissected cells, the vast majority of data in the literature has been obtained in this fashion. The signature of many important components of tumors, such as endothelial, smooth muscle, and inflammatory cells, can be recognized during data analysis, especially if suitable representatives of these cells are included in the database. One can argue that, since the biological properties of a tumor depend on the function of all the various cell types represented in the tumor tissue, information derived from the tissue as a whole actually adds value to the dataset. For example, it may be important to recognize subsets of tumors with higher content of inflammatory cells. It is important to note that, although subtraction *in silico* can provide a reasonable guide to the interpretation of expression patterns, one cannot formally prove that this result is correct without additional experimentation. In general, the difficulties of follow-up studies to verify conclusions drawn from *in silico* subtraction (*in situ* hybridization, immunohistochemistry, or reverse transcription PCR [RT-PCR]) must be weighed against the limitations imposed by cDNA amplification methods. In principle, analysis of microdissected malignant cells will provide a high degree of cell type specificity, but this comes at a considerable cost in terms of specimen processing, as well as carrying the risk of distorting the relative abundance of mRNAs in the amplified product.

After hybridization, a fluorescence image of the microarray is obtained with a scanning device, and the image file is processed with feature extraction software, which converts the raw image to numerical data corresponding to the level of fluorescence in each channel. Commercial instruments and software packages for this purpose perform well. Microarray users must become familiar with the properties of their scanner and use appropriate setting to maximize dynamic range and obtain consistent results between scans. Because it is impossible to use perfectly equivalent amounts of sample in each channel, it is necessary to normalize the sample and reference channels. Two strategies are in wide use, normalization by global intensity or by the use of a set of minimally variable housekeeping genes. The normalized processed data are then output as a spreadsheet for further analysis.

DATA ANALYSIS

Data from a series of tumor samples with expression levels for thousands of genes can present a challenge for analysis. A method for data storage and retrieval in a database is essential. This need not necessarily require the use of enterprise scale databases, but some form of data storage is required. For example, commonly available software, such as FileMaker Pro, can accommodate the needs of many projects.

Although expression profiling studies are sometimes contrasted with traditional hypothesis-driven research, in order to make sense out of microarray data, the researcher must still have a concept of the main questions which it is hoped that a given sample set might answer. Appropriate selection of analysis tools will depend on the questions to be addressed. Certain key questions pervade most cancer-related microarray research: Can two or more types of cancer be discriminated? What genes discriminate them most clearly? Are there genes that discriminate tumors from normal tissues? Are there subsets within tumors of the same apparent class? Are there correlations between expression profiles and other molecular or pathological properties of the tumor? Are there correlations between expression profiles and clinical variables, such as outcome and response to therapy? With what degree of confidence can it be said that these results are not due to chance alone? Do the genes, which arise from these analyses, fall into biologically recognizable pathways? Are these pathways relevant to the tumor phenotype or as potential targets for therapy?

These important questions and the need to develop the mathematical tools to address them have attracted the attention of computer scientists, engineers, and biostatisticians. Numerous computational approaches have been developed, and these will not be reviewed in detail here. However, certain important principles deserve emphasis. The questions listed above vary in difficulty. Some are very easy. For example, finding genes that distinguish colon cancer and glioblastoma will not be a great challenge, and numerous reports support the concept that different cancer types have distinct gene expression profiles *(14,15)*. On the other hand, finding genes that discriminate among stages of colon cancer might be significantly more challenging. There is no reason to be sure, *a priori*, that every question can be answered with confidence by gene expression profiling. For example, the chemosensitivity of a metastatic clone may not be predicted from the gene expression profile of the corresponding primary tumor. As the differences in gene expression narrow between groups of samples, which define clinically relevant groups, the analysis will be less and less forgiving of noise in the data, and the importance of the primary data quality increases. Similarly, larger numbers of samples in each group will be necessary to achieve statistically significant results. Once gene lists are developed, which appear to answer the question posed, they should be used to develop a formal rule-based classifier that might be applied to new samples *(16,17)*. Ideally then, experimental designs should also include a blinded test set, which can be used to validate the results obtained from a "training" set.

Because the number of genes in microarray data sets is always much larger than the number of samples, there will always be some number of genes that appear to differ significantly between groups based on chance alone. There is no method that can prove that this is not the case for any given gene, but probabilistic methods can provide an estimate of the probability that the results are due to chance fluctuations in the data. Alternative methods to address this issue include random permutation tests, leave-one-out analysis, and the introduction of gaussian noise into the data *(18–20)*. These methods help establish whether the data contain an overabundance of genes, which discriminate the samples compared to what would be expected at random. Additionally, it is important to use one of several available methods to rank the genes that discriminate among samples, in order to identify the genes that have the greatest impact on separating groups *(17,18)*. Even with apparently good results, there may be aspects

of the data related to sample selection that will not be apparent until a confirmatory study is attempted. In the end, as in other types of clinical research, there is no substitute for a truly independent confirmatory study.

The methods used to analyze microarray data can be divided broadly into supervised and unsupervised approaches. Clinical correlative studies will utilize supervised methods that divide the samples into groups, for example, responders and nonresponders, according to a known variable, and then search for genes that differ between groups *(21–23)*. Alternatively, microarray datasets present the opportunity for class discovery. This entails the use of unsupervised techniques to search for properties of sample sets that emerge from the data analysis without utilizing the known classification data for the analysis. Unsupervised analysis provides the opportunity to discover unexpected complexities among samples sets. There are a number of excellent examples of the application of this approach to a variety of cancer types *(24–28)*.

VALIDATION OF MICROARRAY DATA

How reliable are microarray data? This question is somewhat laboratory-specific, depending on the precise methodology used. It is also dependent on where a given data point falls on the spectrum of gene abundance. Genes expressed at low levels will not be measured as accurately as more abundant transcripts. In general, when compared to conventional methods, microarray data from experienced laboratories are remarkably accurate *(29)*. Although validation by Northern blot or quantitative PCR methods may be required to confirm or extend important results, inherent data accuracy is usually not a major concern when a pattern of expression is reinforced by a large number of samples. It is somewhat problematic that there may not be an alternative technique that can be used to confirm the expression levels of dozens or hundreds of genes at the same level of accuracy as microarrays, especially when expression levels between sample groups vary by less than two-fold. In this case, the best validation will be obtained from microarray analysis of an additional sample set.

Confirmation at the protein level can also be difficult, since for most genes, a suitable antibody will not be available. Even in the case of genes for which good antibodies capable of staining tissue sections exist, the assay may not have the same dynamic range as hybridization-based methods. In-depth correlation of mRNA and protein expression levels for multiple genes will not be accomplished until accurate quantitative proteomic methods become available.

Tissue microarrays for *in situ* mRNA hybridization or immunohistochemistry provide the possibility of confirmatory studies on large numbers of samples *(30)*. Image analysis of mRNA *in situ* hybridization is remarkably quantitative and agrees well with cDNA microarray data *(31)*. This technology for analyzing a single gene in numerous samples nicely complements the ability of cDNA microarrays to analyze numerous genes in relatively small numbers of samples.

INTERPRETING GENE LISTS

Microarray analysis, whether supervised or unsupervised, ultimately generates lists of genes that discriminate among samples. Making sense of these gene lists presents a significant challenge. Gene names can be misleading, and the majority of genes are linked to little or no functional information. Currently, there are only limited tools that

can automatically parse gene lists into functional categories. Automation of these tasks is an area of active research, but at the present time, only expert perusal of the gene list can optimize gene interpretation in the context of a particular biological question. However, this carries risks. Even if careful statistical methods have been applied to reach this stage of data analysis, there is a real temptation to overinterpret gene lists by drawing tenuous but attractive connections. Most of the interpretations that arise from exploring microarray data are better considered hypotheses than conclusions. It is important to bear in mind that additional forms of experimentation may well be necessary to establish a conclusive connection between a gene and the tumor in which it is expressed.

DISSEMINATING MICROARRAY DATA

The data generated from microarray research far exceeds, in quantity, the limitations set by the usual journal format. Currently, data are provided to the scientific community over the Internet by Web supplements to publications and Web sites maintained by individual laboratories. Public databases have been established at NCBI and European Molecular Biology Laboratory (EMBL), and standards are being developed for the minimal information that should accompany microarray data for publication *(32)*. The availability of significant amounts of data via on-line repositories has led to a number of publications that have reanalyzed existing data, which is a phenomenon that will surely increase dramatically in the future.

CONCLUSION

Although only a brief period has passed since its introduction, cDNA microarray technology has been widely adopted. As more and more studies have confirmed the ability of this technology to contribute to tumor classification and to the recognition of clinically important endpoints, enthusiasm has continued to grow. With time, the quality of microarray data has continued to improve, allowing a progressively higher resolution view of gene expression. This information seems certain to improve tumor classification and, through its impact on therapeutic target discovery, to provide new opportunities for cancer treatment.

REFERENCES

1. Adams, M. D., Kelley, J. M., Gocayne, J. D., et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252,** 1651–1656.
2. Lennon, G. G. and Lehrach, H. (1991) Hybridization analyses of arrayed cDNA libraries. *Trends Genet.* **7,** 314–317.
3. Pietu, G., Alibert, O., Guichard, V., et al. (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* **6,** 492–503.
4. Drmanac, S., Stavropoulos, N. A., Labat, I., et al. (1996) Gene-representing cDNA clusters defined by hybridization of 57,419 clones from infant brain libraries with short oligonucleotide probes. *Genomics* **37,** 29–40.
5. Schuler, G. D., Boguski, M. S., Stewart, E. A., et al. (1996) A gene map of the human genome. *Science* **274,** 540–546.
6. Boguski, M. S. and Schuler, G. D. (1995) ESTablishing a human transcript map. *Nat. Genet.* **10,** 369–371.

7. Pollack, J. R., Perou, C. M., Alizadeh, A. A., et al. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23,** 41–46.
8. Heiskanen, M. A., Bittner, M. L., Chen, Y., et al. (2000) Detection of gene amplification by genomic hybridization to cDNA microarrays. *Cancer Res.* **60,** 799–802.
9. 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,** 467–470.
10. Luo, L., Salunga, R. C., Guo, H., et al. (1999) Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat. Med.* **5,** 117–122.
11. Wang, E., Miller, L. D., Ohnmacht, G. A., Liu, E. T., and Marincola, F. M. (2000) High-fidelity mRNA amplification for gene profiling. *Nat. Biotechnol.* **18,** 457–459.
12. Kitahara, O., Furukawa, Y., Tanaka, T., et al. (2001) Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. *Cancer Res.* **61,** 3544–3549.
13. Alevizos, I., Mahadevappa, M., Zhang, X., et al. (2001) Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. *Oncogene* **20,** 6196–6204.
14. Khan, J., Simon, R., Bittner, M., et al. (1998) Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.* **58,** 5009–5013.
15. Ross, D. T., Scherf, U., Eisen, M. B., et al. (2000) Systematic variation in gene expression patterns in human cancer cell lines [see comments]. *Nat. Genet.* **24,** 227–235.
16. Khan, J., Wei, J. S., Ringner, M., et al. (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* **7,** 673–679.
17. Golub, T. R., Slonim, D. K., Tamayo, P., et al. (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286,** 531–537.
18. Bittner, M., Meltzer, P., Chen, Y., et al. (2000) Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406,** 536–540.
19. Allander, S. V., Nupponen, N. N., Ringner, M., et al. (2001) Gastrointestinal stromal tumors with KIT mutations exhibit a remarkably homogeneous gene expression profile. *Cancer Res.* **61,** 8624–8628.
20. Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M., and Yakhini, Z. (2000) Tissue Classification with Gene Expression Profiles. *J. Comput. Biol.* **7,** 559–583.
21. Gruvberger, S., Ringner, M., Chen, Y., et al. (2001) Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res.* **61,** 5979–5984.
22. Korn, E. L., McShane, L. M., Troendle, J. F., Rosenwald, A., and Simon, R. (2002) Identifying pre-post chemotherapy differences in gene expression in breast tumours: a statistical method appropriate for this aim. *Br. J. Cancer* **86,** 1093–1096.
23. van ‘t Veer, L. J., Dai, H., van de Vijver, M. J., et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536.
24. Perou, C. M., Jeffrey, S. S., van de Rijn, M., et al. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* **96,** 9212–9217.
25. Perou, C. M., Sorlie, T., Eisen, M. B., et al. (2000) Molecular portraits of human breast tumours. *Nature* **406,** 747–752.
26. Alizadeh, A. A., Eisen, M. B., Davis, R. E., et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling [see comments]. *Nature* **403,** 503–511.
27. Garber, M. E., Troyanskaya, O. G., Schluens, K., et al. (2001) Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. USA* **98,** 13,784–13,789.
28. Bhattacharjee, A., Richards, W. G., Staunton, J., et al. (2001) Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. USA* **98,** 13,790–13,795.

29. Amundson, S. A., Bittner, M., Chen, Y., Trent, J., Meltzer, P., and Fornace, A. J., Jr. (1999) Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. *Oncogene* **18,** 3666–3672.
30. Kononen, J., Bubendorf, L., Kallioniemi, A., et al. (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* **4,** 844–847.
31. Mousses, S., Bubendorf, L., Wagner, U., et al. (2002) Clinical validation of candidate genes associated with prostate cancer progression in the CWR22 model system using tissue microarrays. *Cancer Res.* **62,** 1256–1260.
32. Brazma, A., Hingamp, P., Quackenbush, J., et al. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet.* **29,** 365–371.