ARTHRITIS & RHEUMATISM
Vol. 46, No. 4, April 2002, pp 859–861
DOI 10.1002/art.10236

EDITORIAL

DNA Microarrays: Boundless Technology or Bound by Technology? Guidelines for Studies Using Microarray Technology

Gary S. Firestein[1] and David S. Pisetsky[2]

The biotechnology revolution has led to unprecedented progress in elucidating the structure of the human genome and determining patterns of gene expression relevant to disease pathogenesis. Each new breakthrough produces an immediate and often intense flurry of activity as investigators explore its place in the scientific repertoire and determine the insights it provides for understanding disease. Polymerase chain reaction (PCR) is an example of such a transforming technology which, in a short time, facilitated a quantum leap in our ability to clone novel genes and quantify gene expression at the level of messenger RNA (mRNA). As with all "hot" new methods, many years passed before the strengths and weaknesses of PCR were recognized and more powerful applications, such as real-time PCR, were developed and refined.

Microarray technology represents another cutting-edge technology that promises extraordinary advances in the study of disease. As described in an article by Grant et al in this issue of *Arthritis & Rheumatism* (1), microarray technology provides scientists the tools to scan simultaneously the array of expressed genes in a cell and glean myrid information about cellular function. In the face of such power and precision, it may appear as if hypotheses are superfluous to the pursuit of research and that specific questions need no longer be asked. Indeed, microarray technology, buttressed by elaborate computer software of great sophistication, can give the illusion that all of the information can be "downloaded" from the cell and its secrets revealed to the inquisitive mind. The siren song of microarrays is indeed strong and tantalizing.

Of course, research is never quite that simple. With every new technology, the siren song gives way to harsh reality. Experience with a technology is often sobering, reminding investigators of the complexity of life processes as well as scientific investigation. In the case of microarrays, this complexity comes with an unprecedented amount of data that variously seems to be a treasure chest, a morass, and an assault. Like all emerging technologies, microarrays necessitate great respect and caution lest the resulting data be misused if not abused.

In the last year, *Arthritis & Rheumatism* has received a large number of manuscripts describing studies using microarray technology. These manuscripts address the full gamut of rheumatologic disease and involve material from patients, animal models, and in vitro cell models. By their nature, these studies, which entail a field in its infancy, raise difficult questions concerning the criteria for judging the technical aspects of the research as well as its scientific validity.

Because of the confusion that faces the investigator, reviewer, and reader confronted with microarray data, we have long pondered how best to promote the publication of articles describing studies using this technology. As a result of extensive discussion among members of the editorial board of *Arthritis & Rheumatism*, we have developed suggested guidelines for the submission and review of reports of investigations involving microarray technology. These guidelines acknowledge the strengths and limitations of this technology and indicate the rigor we believe necessary to publish reports of microarray-based studies in our journal. The guidelines are tentative and should be viewed as a work in progress that will be revisited as the field progresses. Nevertheless, with these guidelines, we hope to provide the scientific community with broad criteria that investigators can use in the design of their studies and, in turn,

[1]Gary S. Firestein, MD: University of California, San Diego School of Medicine, La Jolla, Deputy Editor, *Arthritis & Rheumatism*; [2]David S. Pisetsky, MD, PhD: Durham VAMC and Duke University Medical Center, Durham, North Carolina, Editor, *Arthritis & Rheumatism*.

Address correspondence and reprint requests to David S. Pisetsky, MD, PhD, Durham VA Medical Center, 151G, 508 Fulton Street, Durham, NC 27705.

Submitted for publication December 19, 2001; accepted December 20, 2001.

Table 1. Criteria for evaluating microarray studies

1. Reproducibility must be demonstrated, including rigorous evaluation of the run-to-run variability for each gene.
2. Detailed statistical analysis is required, including appropriate corrections for repeated or multiple measurements.
3. Homogeneous cell populations should optimally be studied, to reduce the complexity of analysis.
4. A non-array method must confirm changes in the expression of key genes.

reviewers and readers can use in their evaluation. Table 1 highlights the issues.

Issue 1: Reproducibility of the methods

As with any technique, reports of microarray studies should document the accuracy and precision of the data, including evaluation of the run-to-run variability. This documentation is critical for determining sample size and establishing the power for a study that will subsequently be subjected to rigorous statistical analysis. Arbitrary setting of thresholds for "significant changes" (e.g., 2-fold increase in a particular gene) is not supported by current scientific information, and this practice should be discouraged unless appropriate validation can be provided. The variability for each individual gene should be calculated, thereby permitting appropriate adjustments to avoid incorrectly identifying an observed "increase" as significant (i.e., the "false discovery rate") (2). Unless these statistical corrections are applied, the false discovery rate can vary from 60% to 80% of identified genes even if the threshold is set very high (even 4-fold over control). With use of well-defined statistical algorithms based on the inter-experiment variability, this rate can be decreased to ~10%.

Issue 2: Statistical analysis

The statistical analysis of data is perhaps the most difficult problem associated with the use of microarray techniques. Statistical evaluation is absolutely necessary to support claims of an increase or decrease in gene expression with arrays or any other such method. Such rigor requires multiple experiments and analysis by standard statistical instruments. For instance, as mentioned above, it is not sufficient to say that the expression of a particular gene is 2-fold greater in a sample compared with control and that this difference is statistically significant based on historic controls for reproducibility. Each study using microarray technology should include a sufficient number of independent experiments to allow analysis of the results by methods similar to those used for any other data set (e.g., determination of the mean ± SD with a specific *P* value for the comparison). The number of arrays and replicate experiments needed to achieve statistical significance is dependent on an accurate solution to issue 1 above, i.e., the coefficient of variation. While we recognize that the performance of a sufficient number of experiments can entail significant expense, thrift cannot overrule the need for scientific validity.

Equally important, appropriate statistical corrections must be applied to account for multiple comparisons. For example, comparing microarrays with up to 30,000 data points will virtually always lead to apparently "significant" differences in perhaps hundreds of genes. Corrections for repeated or multiple measurements (e.g., Bonferroni correction) are essential to avoid pitfalls of ascribing significance to chance events.

An adjunct method to validate a possible difference between 2 sets of microarray data involves "shuffling" the data (2). For example, an analysis of 2 sample sets could demonstrate that the expression of 200 genes, of 30,000 genes measured, "differs" between the 2 groups. To determine if the expression of these genes is truly different, the data can be reanalyzed after systematically relabeling individual members of the sample set. If, for example, one is comparing gene expression in 10 osteoarthritis (OA) and 10 rheumatoid arthritis (RA) synoviocyte lines, one could shuffle an OA and an RA line so that they are "inadvertently" included in the wrong group for the statistical analysis. This process can be repeated until every combination has been tested. If the pattern of gene expression between the 2 groups truly differs, then the number of genes that are "different" should decrease as the shuffling becomes increasingly random. In our experience, "statistically significant" differences identified between 2 data sets can still be observed (albeit with different genes) even when the assignment to a particular sample group is entirely random. Under these circumstances, it is not possible to conclude that the originally observed differences were significant.

Issue 3: Sample heterogeneity

The hazards in interpreting microarray data are greatly amplified when the technique is applied to heterogeneous cell populations. Even with a uniform cell population, it can be difficult to use this technique to validate the significance of an increase or decrease in gene expression. However, the challenges of analyzing heterogeneous tissue, such as synovium or whole limb, are huge, if not insurmountable. What do we learn when the expression of a particular gene is increased in this

type of sample? Is this increase in the level of gene expression due solely to differences in the composition of the cell population or is it related to a cellular event that is critical to disease pathogenesis? For example, does an increase in the number of T cells in an arthritic joint cause an increase in some genes (e.g., T cell receptor) that may be only incidentally related to the inflammatory process? How does one weight putatively pathogenetic genes, such as interferon-γ, to correct for the bias created by changes in cell number? This problem is exacerbated further in studies using whole tissue extracts, where the gene with altered expression could be found in bone, cartilage, or synovium (or even skin?).

In principle, we have no objection to the use of whole joint extracts or heterogeneous samples with some techniques, such as Western blot or Northern blot analysis, in hypothesis-driven studies. The issue of data interpretation, however, can become overwhelming when using a method such as microarray technology, where the sheer mass of information poses inherent analytic problems. Because of these questions, this journal strongly recommends that investigators focus studies on homogeneous cell populations until the other methodologic and data analysis problems can be resolved.

Issue 4: Independent confirmation

Because of the statistical issues raised by microarray technology, it is very important that the findings be confirmed using an independent method, preferably with separate samples rather than retesting of the original mRNA. Because data resulting from a microarray are so extensive (i.e., changes in the expression of tens or hundreds of genes), it is impossible to retest all of the data. Nevertheless, it is incumbent upon investigators to evaluate a reasonable number of genes.

"Confirmation" is not without its own problems. For example, a study could demonstrate a 4-fold increase in expression of a particular gene with the use of microarray but a much greater (or lesser) change with another method, such as Northern blot analysis. Does such a result actually confirm the microarray findings, or does a major quantitative difference from results obtained with a "gold standard" technique raise new questions about the validity of the microarray data? The problem here lies in the fact that the fold increase in individual genes observed with the use of microarrays can vary considerably from that found with "gold standard" methods; in some cases, the differences in findings between the 2 methods can be much greater than the original change reported for microarray alone. Under these circumstances, the reliability of the chip data is far less certain. In our view, demonstrating that gene expression moves in the same direction may not necessarily represent sufficient confirmation if the fold changes observed using independent methods are not reasonably similar. While raising the bar for confirming the results of microarray data, these considerations emphasize the preliminary nature of the chip data and help define the purpose of microarrays as a screening test to identify the best targets for hypothesis-driven studies.

What place will microarray technology hold in our scientific armamentarium? This is a rapidly changing field, and there is still no consensus about the ultimate utility of this technology in probing disease mechanisms and identifying targets for new treatments. While application of the guidelines outlined above will not resolve the formidable issues involved in interpreting microarray data, they do offer a framework for the conduct and evaluation of research during this time of rapid technological advance. Data showing differences in gene expression between heterogeneous populations, collected using current technology, may be of uncertain value. These studies, however, can narrow the search for disease-related genes and serve as a springboard for subsequent hypothesis-driven work.

Microarray technology will undoubtedly become one of the most intensively utilized methodologies in the coming years and likely will lead to some unique insights and dazzling successes. In other disciplines, the true power of the chips has been amply demonstrated when they have been used to characterize large sample sets of homogeneous cell populations. An example of such success is the use of microarrays to predict treatment response in patients with B cell lymphomas. In those studies, characterization of patterns of gene expression by microarray techniques allowed the segregation of lymphomas into 2 broad categories with differences in response to therapy (3). Applying this approach to the study of arthritis and related diseases that involve multiple cell types, however, will require considerable care in sample selection and data interpretation. Until the technical and analytic issues can be resolved, the guidelines described herein represent a reasonable step in harnessing the power of technology while protecting the literature from a mass of data that may be not only confusing but misleading.

REFERENCES

1. Grant EP, Pickard MD, Briskin MJ, Gutierrez-Ramos J-C. Gene expression profiles: creating new perspectives in arthritis research. Arthritis Rheum 2002;46:874–884.
2. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116–21.
3. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–11.