The promise and perils of microarray analysis

William H. Catherino, MD, PhD,[a,b,*] **Phyllis C. Leppert, MD, PhD,**[a,b]
James H. Segars, MD[a,b]

Department of Obstetrics and Gynecology, Uniformed Services University of the Health Sciences[a]*; Reproductive
Biology and Medicine Branch, National Institute of Child Health and Human Development, National Institutes
of Health,*[b] *Bethesda, MD*

Microarray analysis has provided a novel means of identifying clues into the mechanisms of disease development. As a methodology, microarray analysis holds the promise for genome-wide screening in which 2 tissues (diseased and normal) are compared, and molecular pathways that defined the phenotype of the disease could be precisely defined. Alternatively, microarray experiments can be used to differentially compare pathologically similar diseased tissues to predict response to chemotherapy and risk of recurrence. However, the clinician should be aware that various sources of error can influence microarray analysis results. Sources of error can be minimized but not eliminated, explaining why meticulously conducted experiments in different laboratories or using different platforms result in different lists of genes. Confirmation and validation of genome-wide microarray results using ancillary methods remains a critical step. With proper confirmatory studies and cautious interpretation, microarray analysis represents a powerful tool for molecular discovery.
© 2006 Mosby, Inc. All rights reserved.

The inception of microarray analysis, followed closely by the completion of the human genome project and the availability of genome-wide chips, resulted in the promise of testing experimental effects across the entire human genome. Rather than going through the laborious task of identifying gene products one-by-one that were individually associated with specific disease processes, it became possible to compare normal and abnormal tissue to identify patterns of gene expression that might define the phenotype of the disease of interest. Such patterns might be unique to the disease, possibly leading to identification of novel therapies.

Although potential application of microarray technology to the clinical setting is enticing, there are important limitations to consider. Obviously, not all diseases are associated with or defined by differences in messenger RNA (mRNA) levels, and microarrays will not necessarily shed light on such conditions. In addition, the potential of the method is affected by sources of error or variability inherent to the methodology. While a thorough discussion of the bioinformatics and analysis

* Reprint Requests: William H. Catherino, MD, PhD, Department of Obstetrics and Gynecology, Uniformed Services University of the Health Sciences, Building A, Room 3078, 4301 Jones Bridge Rd, Bethesda, MD 20892.

E-mail: catheriw@mail.nih.gov

of microarray data is clearly beyond the needs of the clinician and scope of the *Journal*, it is useful to briefly consider the nature of information provided by microarrays. First, there are several types of arrays, or platforms, that may be used: custom complementary DNA (cDNA) arrays containing large portions of transcripts or commercially available arrays such as those offered by Applied Biosystems (www.appliedbiosystems.com), or Affymetrix Corporation (www.affymetrix.com), that are composed of oligonucleotide sequences.

For purposes of illustration, let's consider one of the best characterized and validated commercial sources of microarrays, those offered by Affymetrix Corporation, a platform we have used in our research.[1] Each gene product is represented on Affymetrix GeneChips by 11 pairs of short oligonucleotides (25 bases) of correct (perfect match [PM]) and corresponding mismatch (control [MM]) sequences. The oligonucleotides are synthesized in situ on a chip. To perform the array, RNA is labeled (the target) and hybridized to the chip containing the oligonucleotides and the differences in hybridization between correct (PM) and control (MM) sequences are then used to control for erroneous signal (for more detailed explanation, see "Data Analysis Fundamentals" on the Affymetrix Web site). To control for nonspecific hybridization, MM sequences with a single base pair (bp) change in the middle of the expected nucleotide sequence are also included. By taking into account the likelihood of repeated hybridizations and subtracting the nonspecific hybridization, the relative expression can be determined.[2,3]

Although the theoretical likelihood of false detection of a single gene is low, the experimentally observed, practical error rate can be much higher. Contrary to expectation, control MM sequences may have stronger signal than correct (PM) sequence for gene products because of cross-hybridization.[4] In addition, there are indeterminate values, false-positives, false-negatives, and incorrect assignment of probes. It is important to realize that microarray results are referenced against the gene products derived from the human genome project, although reassigned probe sets have been made available[4] (http://mriweb.moffitt.usf.edu/mpv/). Some investigators have found that only about 70% of PM probes are responsive to difference in target (RNA) concentration; 10% are unresponsive and 20% are invariant (www.expressionanalysis.com). Practically, this means that unless specific measures are taken, experimental error with this platform can be 20% or higher.[5] It should be emphasized that Affymetrix GeneChips have been reported to have the best within-platform reproducibility among the commercially available microarray chips[6] and the issue of variability is found across all platforms. For some experimental comparisons, an error rate of that magnitude may be acceptable.

To illustrate the problem of variability in microarray experiments, we tested 2 RNA aliquots from the same leiomyoma and performed microarray analysis using the Affymetrix HG-U133 A and B chips. The samples were handled by the same technician in the same core facility with the same equipment, but were performed on different days. The RNA sample integrity was confirmed by gel electrophoresis, and each sample demonstrated a 260/280 spectrophotometric ratio of at least 1.8. Normalization was performed. Some representative data are presented in Tables I and II. Despite the fact that the RNA samples were presumably identical, 9.9% of the genes displayed at least 2.5-fold overexpression in the second microarray analysis compared with the first, whereas 4.4% displayed at least 2.5-fold underexpression. For the purposes of this experiment, we used a cutoff of 250 pixel-intensity units as a lower boundary, and 60,000 pixel-intensity units as an upper limit. These limits are stricter that those described by Tadesse and Ibrahim.[7] Of course, most investigators repeat experiments several times before results are published, something we did not do with this simple example, and it is certainly possible that with replication the results might improve. However, replication serves only to minimize error inherent in the microarray technology. Even when the patient and laboratory were held constant, there was still significant "noise" in the experiment. Given the number of genes involved, a small amount of error can translate into a large number of genes incorrectly assumed to be differentially expressed. Again, the issue is not a problem particular to Affymetrix Gene-Chips, but may also be due to the large number of factors that can affect reproducibility of results. Other laboratories have found similar reproducibility issues.[6]

Given this potential variability, and because meticulously conducted experiments in different laboratories with different platforms can uncover different sets of genes,[8] how is the clinician to know which microarray data (lists) are correct or and which genes biologically relevant? The skeptic would say: lists alone are not very meaningful and one should only consider results validated if the difference in gene expression is verified by a laboratory-based method such as real-time reverse transcriptase polymerase chain reaction (RT-PCR). For a few genes this is practical, but it is impractical to evaluate the entire genome with real-time RT-PCR to exclude false-negative results, for instance. In silico validation is another option.[9] The microarray proponent could suggest that for any particular experiment there is a correct analysis and if conducted properly the results will be reliable. This could theoretically be true in genetically defined populations, but the clinician is interested in consistent findings in diseases in an out bred species and is rightly concerned that different experiments on the same disease using the same approach often yield different lists of genes.