Microarrays

Applications and Pitfalls

Hassan M. Fathallah-Shaykh, MD

Microarrays are simple assays that measure the relative expression levels of tens of thousands of genes. Excitement about their importance and potential contributions to biology and medicine has been intense. Nonetheless, recent insights into the limitations and pitfalls of microarrays have led to caution about data interpretation. Microarrays are very useful but they are also very misleading; better data analysis tools are needed to improve accuracy.

Over the past half century, scientists have studied cause-and-effect relationships between known genes and biological phenotypes or human disease. Recent technological advances have changed the landscape of biomedical research. The complete genomes of several organisms are now available, and the expression of tens of thousands of genes may be assayed by microarrays. Genomes are rich sources of complex genetic information, most of which is unknown and unpredictable. Hence, the term discovery has been introduced to imply finding without preconceived bias which genes are relevant to a biological phenotype and how the genes interact.

In a single assay, microarrays generate tens of thousands of measurements of the relative levels of messenger RNA expression. When first developed, microarrays appeared to hold great promise for translating genomics into significant advances in basic biology and medicine. The National Institutes of Health (Bethesda, Md), universities, and drug companies have invested heavily in various applications of microarrays. Nonetheless, recent findings have uncovered major pitfalls that cast doubt on the interpretation of microarray data. Herein, I review the technology of complementary DNA (cDNA) microarrays, their applications and pitfalls, and future directions in data analysis.

THE EXPERIMENTAL SYSTEM

Spotted arrays may include tens of thousands of cDNAs laid on glass slides. Each experiment uses 2 RNA samples and measures the relative expression level of the cDNAs in 1 messenger RNA as compared with the other (Figure 1). The messenger RNAs are reverse transcribed to cDNAs and labeled with fluorescent dyes, mixed, and hybridized to the glass slide. After washing, the spot-bound fluorescent dyes are excited by lasers of appropriate wavelengths to generate 2 "scanned" images, which correspond to the samples. Images are analyzed to quantify (1) the signal within each spot and (2) a small rim of background surrounding each spot. The principal measurement is the expression ratio of each spot:

\[
\log_2 \left(\frac{\text{Background-subtracted Intensity in Sample}}{\text{Background-subtracted Intensity in Reference}}\right)
\]

A log2(ratio) greater than 0 implies up-regulation and a log2(ratio) less than 0 implies down-regulation. A data set of a single experiment contains tens of thousands of ratios.

Microarray experiments using spotted arrays are usually designed to compare each of several samples to a single reference RNA that is common to all experiments. The data are expressed in a ma-

Author Affiliations: Department of Neurological Sciences, Section of Neuro-oncology, Rush University Medical Center, Chicago, Ill.
tions of central nervous system embryonal tumors, glioblastomas, large B-cell lymphoma, and breast carcinoma.1,10-15 Preliminary proofs of principles include reports of pattern discovery based on molecular expression that predict new classifications of disease. Specifically, classifications of the complementary DNAs (cDNAs) in the sample vs the reference. mRNA indicates messenger RNA.

The enthusiasm about the potential of microarrays has been intense.7-9 Experimental designs are usually aimed at discovering (1) patterns of expression that classify disease phenotypes and predict clinical behavior or (2) molecular targets and systems that create the biology. The first goal is based on the intuitive idea that genome-scale molecular expression refines the pathological classification of disease. Specifically, classifications based on molecular expression are expected to be more accurate and sensitive than those based on microscopy. Preliminary proofs of principles include reports of patterns of genetic expression that predict new classifications of central nervous system embryonal tumors, gliomas, large B-cell lymphoma, and breast carcinoma.1,10-15 For example, the molecular classes may either replicate the pathological distinction or divide the subjects within the same pathological class into subgroups that predict distinct clinical behaviors like long-term vs short-term survival times and drug response vs resistance.

The idea that the global transcriptional response constitutes molecular phenotypes has recently received attention.12,16,17 In this model, phenotypes are created by molecular systems in which single genes or molecules belong to rich networks of dynamic molecular interactions that include transcriptional regulation, signaling pathways, protein-protein, and protein–nucleic acid interactions.16,18 Examples of microarray applications in systems biology include the discovery of (1) the regulation of the transcriptional response when yeast cells encounter nutrients, (2) the yeast galactose-utilization pathway, and (3) the principles of balanced genetic expression and opposing molecular functions behind the phenotypes of meningiomas and cultured gliomas.16,19-22 Theoretically, one could apply microarrays to discover new molecular classifications of neurological diseases, to study and define the molecular systems that create each individual phenotype, and to perturb the network to find the best targets that transition the whole system between phenotypes.

POTENTIAL APPLICATIONS OF MICROARRAYS IN BIOLOGY AND MEDICINE

The enthusiasm about the potential of microarrays has been intense.7-9 Experimental designs are usually aimed at discovering (1) patterns of expression that classify disease phenotypes and predict clinical behavior or (2) molecular targets and systems that create the biology. The first goal is based on the intuitive idea that genome-scale molecular expression refines the pathological classification of disease. Specifically, classifications based on molecular expression are expected to be more accurate and sensitive than those based on microscopy. Preliminary proofs of principles include reports of patterns of genetic expression that predict new classifications of central nervous system embryonal tumors, gliomas, large B-cell lymphoma, and breast carcinoma.1,10-15 For example, the molecular classes may either replicate the pathological distinction or divide the subjects within the same pathological class into subgroups that predict distinct clinical behaviors like long-term vs short-term survival times and drug response vs resistance.

The idea that the global transcriptional response constitutes molecular phenotypes has recently received attention.12,16,17 In this model, phenotypes are created by molecular systems in which single genes or molecules belong to rich networks of dynamic molecular interactions that include transcriptional regulation, signaling pathways, protein-protein, and protein–nucleic acid interactions.16,18 Examples of microarray applications in systems biology include the discovery of (1) the regulation of the transcriptional response when yeast cells encounter nutrients, (2) the yeast galactose-utilization pathway, and (3) the principles of balanced genetic expression and opposing molecular functions behind the phenotypes of meningiomas and cultured gliomas.16,19-22 Theoretically, one could apply microarrays to discover new molecular classifications of neurological diseases, to study and define the molecular systems that create each individual phenotype, and to perturb the network to find the best targets that transition the whole system between phenotypes.

PITFALLS OF MICROARRAYS

Following the initial hype and excitement about microarrays, their pitfalls and limitations are causing a hard reality check. Current methods for microarray expression data analysis require numerous samples and yield measurements of low specificity. Kothapalli et al23 examined microarray data from 2 different systems. They report inconsistencies in sequence fidelity of the spotted microarrays, variability of differential expression, low specificity of cDNA probes, discrepancy in fold-change calculations, and lack of probe specificity for different isoforms of a gene. Ntzani and Ioannidis24 examined 84 large-scale microarray expression data sets that address major clinical outcomes including death, metastasis, recurrence, and response to therapy. They found that these studies show variable prognostic performance. Tan et al25 examined gene expression measurements generated from identical RNA preparations that were obtained using 3 commercially available microarray platforms from Affymetrix, Amersham, and Agilent. Correlations in gene expression levels and comparisons for significant gene expression changes in this subset showed considerable divergence across the different platforms. Michiels et al26 reanalyzed data from the 7 largest published studies that have attempted to predict prognosis of patients with cancer on the basis of DNA microarray analysis. The results reveal that the list of genes identified as predictors of prognosis was highly unstable and molecular signatures were strongly dependent on the selection of patients in the training sets. In addition, 5 of the 7 studies did not classify patients better than chance. The poor specificity and reproducibility are not surprising considering all the experimental variables that affect the quality of the data sets. These include variations in the laboratories, individuals, probe labeling, biochemical reactions, scanners, and lasers. Because of the low specificity, validation by other methods for measuring gene expression has become the “gold standard.”25,27 However, biological samples are not always abundant, and the price tag of validating all the genes discovered by microarray expression profiling is astronomical.

Figure 1. A schematic portraying expression profiling of a sample vs a reference by spotted microarrays using probe-switching (dye swap) experiments. The results yield replicate expression levels of the ratios of the complementary DNAs (cDNAs) in the sample vs the reference. mRNA indicates messenger RNA.
THE NATURE OF THE PROBLEM

The specificity of the discovery should be stringent when the data sets consist of tens of thousands of genes and contain a predominant majority of noise. To illustrate, let us consider the example of a data set containing 500 true states of genetic expression (up-regulated or down-regulated) and 19,500 false-positive states (Table). Specificities of 99% and 95% yield 195 and 975 false-positive expression states, respectively. Thus, an analytical method having 100% specificity and 99% specificity discovers 695 genes (500 + 195), 28% (195/695) of which are false positive. Another method having 50% sensitivity and 99% specificity yields 445 genes (250 + 195), 44% (195/445) of which are false positive. This example illustrates the limitations of statistical significance when noise is predominant.

Microarrays assay for the relative expression levels of a cDNA (1) in a biological sample as compared with another (2) relative to other cDNAs within the same sample. The accuracy of fold changes is critical for data analysis. The results of Kothapalli et al. reveal poor reproducibility and discrepancies of fold-change calculations between microarrays (interarray). Furthermore, the accuracy of calculations of fold changes of genes within a single microarray (intra-array) is not known. Low specificity, the preponderance and heterogeneity of noise, and inaccurate fold-change calculations impose significant limitations on data analysis. For example, apparent molecular classifications may be caused by data set-specific noise and the results of 1 laboratory may disintegrate when tested independently. Furthermore, variations in gene expression levels between biological samples may be caused by noise and not biological heterogeneity.

HIGHLY SPECIFIC EXPRESSION DISCOVERY

Recent reports describe mathematical models that shed light on the behavior of noise in microarray data sets and algorithms that discover highly specific states of genetic expression (up-regulated or down-regulated) from genomewide expression profiling. The mathematical models incorporate the principles of (1) preponderance and (2) heterogeneity of noise. The preponderance of noise implies that (1) the overwhelming majority of the genes on the array are not differentially expressed between samples (true negatives) and (2) the truly negative genes generate false-positive expression data (noise). Noise heterogeneity implies that the distribution of noise varies between data sets depending on quality. These principles may be summarized as follows:

1. Each sample vs reference comparison generates tens of thousands of expression ratios.
2. The model is based on the idea that less than 5% of all the genomic genes are truly differentially expressed between the sample and reference (true positives). The expression levels of the other more than 95% are not expected to be different (true negatives).
3. Even when the expression levels of the genes do not differ between the sample and reference, the predominant majority of their measured expression ratios are not equal to 1 (noise, artifacts, or false positives).
4. The distributions of the false positives vary widely between experiments; the variability is determined by quality.
5. True-positive (<5%) and false-positive ratios (>95%) share the same distributions.

The mathematical tools generate highly specific discovery by modeling and filtering noise (Figure 2). The use of mathematical modeling and filters is common; to name a few examples, engineers apply filters to solve problems of noise in cellular telephones, digital music, and digital television.

SIGNIFICANCE AND FUTURE DIRECTIONS

Highly specific genome-scale discovery of states of genetic expression has applications in all aspects of biology and medicine; it facilitates hypothesis-driven research and sets the stage for studies in systems biology. Several models that explain the relationship of genotype to phe-
notype have evolved over the past 40 years. First is the model of a single genetic lesion causing a phenotype; an example is sickle cell disease. A second model is that of several genotypes causing the same phenotype; examples include malignant brain tumors and Alzheimer disease. A third model is that of a single genetic lesion causing distinct phenotypes depending on polymorphisms; examples include hereditary Creutzfeldt-Jakob disease and fatal familial insomnia. Data from the highly specific genome-scale discovery in meningiomas are consistent with a fourth model of complex molecular systems.16 These 4 models are not exclusive; for instance, complex molecular systems biology, but they can also be very misleading. Not only many fields in physics, the full potential of microarrays awaits advances in mathematics. We ought to step back to the drawing board to develop better tools for data analysis.

Accepted for Publication: July 29, 2004.
Correspondence: Hassan M. Fathallah-Shaykh, MD, Department of Neurological Sciences, Section of Neurooncology, Rush University Medical Center, Chicago, IL 60612 (hfathall@rush.edu).

REFERENCES
32. Li JL, Hayden MR, Almqvist EW, et al. A genome scan for modifiers of age at onset of familial insomnia.29 Data from the highly specific genomic discovery in meningiomas are consistent with a fourth model of complex molecular systems.16,18,30,31 In this model, single genes or molecules of the cell belong to rich networks of molecular interactions that include transcriptional regulation, signaling pathways, protein-protein, and protein–nucleic acid interactions.16 These 4 models are not exclusive; for instance, complex molecular systems biology, but they can also be very misleading. Not only many fields in physics, the full potential of microarrays awaits advances in mathematics. We ought to step back to the drawing board to develop better tools for data analysis.

Accepted for Publication: July 29, 2004.
Correspondence: Hassan M. Fathallah-Shaykh, MD, Department of Neurological Sciences, Section of Neurooncology, Rush University Medical Center, Chicago, IL 60612 (hfathall@rush.edu).

©2005 American Medical Association. All rights reserved.