Is it on or off?

The Use of Microarrays in Functional Genomics

Mark Hsu
Biochemistry 118
Professor Doug Brutlag
5/29/02

At a time when the human genome has nearly been sequenced and the genomes of various organisms have already been completed, there is a plethora of information waiting to be uncovered in the sequences that make up these genes. Despite such initial breakthroughs in terms of discovering the genetic, structural makeup of organisms, a formidable task lies ahead: to be able to determine the functions of all these genetic sequences. As Mark Schena of Stanford University observes, "Understanding human behavior, disease and other health issues thus requires more than just a knowledge of genes and genomes: one must understand the cellular, physiological, cultural and ecological context in which genomic instructions are being read (1)." Indeed, the billions of DNA bases alone do not directly tell us about the function of genes, cellular processes, and mechanisms of disease.

The advent of genomic technologies such as RNA Northern blots, ribonuclease protection assays, and polymerase chain reaction (PCR) has only begun to explore the surface of interpreting this vast sequence information by identifying genes and evaluating gene expression. An inherent limitation to these methods, though, is that these types of genetic studies are only able to investigate the role of a few genes at a time, and as a result, preclude an overall genetic survey. These tests are also sequential in nature, and thus require a great deal of time and work to be able to evaluate a small number of genes at once (2).

Microarrays have emerged as an extremely powerful tool that bioinformatics has to offer in analyzing the genome, combining the insight that experimental methods and biological processes can potentially reveal with the computational methods and analytical precision of computers. First introduced in 1996, microarrays promise to serve as a

widespread tool in greatly accelerating scientific research and revolutionizing medical diagnosis. Based on complementary DNA (cDNA), these arrays work by hybridizing labeled RNA or DNA to DNA molecules attached to specific locations on the glass surface of the chip, which can then be detected through fluorescence by a computer scanner. Microarrays promise to play a vital role in functional genomics by interpreting the genome to help us understand how genes and their gene functions are integrated to give rise to functional cells and organisms.

A typical gene expression analysis with the DNA microarray could be used in the following procedure by comparing two samples of mRNA. For example, mRNA could be collected from liver cells that received a pharmaceutical drug and be contrasted to an mRNA sample taken from control liver cells that did not receive any treatment. cDNA is first prepared from each mRNA sample by using fluorescently labeled nucleotide precursors. Then, the two cDNA samples can be labeled with differently colored fluorescent labels to be able to differentiate between the two setups. Finally, by adding the labeled cDNAs to the chip, binding occurs when cDNA sequences hybridize specifically to complementary sequences of bases on the microarray; each gene represented by a spot of DNA. Running the chip through a scanner, the computer can then calculate the ratio of cDNA binding of each sample at each spot. Thus, the relative abundance of gene expression can be determined, and by looking at which genes were activated, comparisons can be made to pre-existing knowledge of the activated genes to assess the extent of drug efficacy or potential side effects (3).

Microarrays are very versatile instruments whose results can be interpreted in a variety of different scientific contexts and fields. Three primary areas where microarrays

are expected to make a significant impact are in expression profiling, drug development, and medical diagnosis.

One of the most important applications being employed with microarrays is with a phenomenon called expression profiling, which measures the amount of gene expression by assessing the levels of mRNA abundance. For the most part, the assumption is made that changes in mRNA amounts correspond to changes in the amount of protein made. A clear advantage to the usage of these assays is that they focus on the "expressed" segments of the genome; especially since the coding sequences account for only about 3% of the genome, gene expression analysis can simplify the human genome's complexity by up to thirtyfold (1). The application of expression profiling is invaluable because once scientists pinpoint which proteins are upregulated following tissue exposure to different conditions, they can then surmise how the tissue deals with such modifications of conditions. This approach of expression profiling brings the genome to life: microarrays can monitor the multi-gene patterns of expression and provide mechanistic insight as to what regulatory mechanisms, cellular functions, and biochemical pathways are involved in one's various physiological processes. Since multiple genes are often turned on or off during a response to a stimulus such as a drug or infection, the cellular response constituted by the multitude of genes most likely operates via the same pathway. Likewise, one can then tell us how the failure of particular genes to compensate for these changes can result in disease development. Correlations between the presence of certain expression markers and disease susceptibility can be drawn, as altered expression patterns often serve as a signature of human diseases (4).

Consequently, microarrays can then be used to identify genes and proteins that can be implicated in a disease-causing pathways by studying differential gene expressions. The use of microarrays in drug development bypasses two major barriers of traditional drug development: 1) the number of potential targets was dictated by the number of cloned genes; 2) drug targeting required a thorough understanding of the disease's pathophysiology. (5). Instead, the simple up- or downregulation of gene activity can either signify the cause of pathophysiology or the result of the disease; consequently, targeting these pathological gene products can not only address the disease itself, but also alleviate symptoms. Once such genes and proteins have been isolated, they can accordingly serve as potential targets for drug development in therapeutic intervention (5). Another benefit that scientists can use with microarrays is to eliminate potential drug candidates that produce intolerable side effects. For example, if scientists want to eliminate toxic drug candidates for cardiovascular disease, they can first build up an expression profile for heart cells by documenting the gene responses of these cells to pre-existing toxic drugs and chemicals. Then, by testing the drug candidates through the microarrays, these results can be compared to the expression profiles of pre-existing compounds to see if there are any comparable gene responses. If so, scientists can conclude that the drug candidate carries similar effects as the compound of the expression profile (6).

On the medical front, physicians are looking to employ microarrays as a powerful diagnostic tool that would enable them to tailor personalized treatment plans for their patients, based on the patients' genomes. As has been known for a very long time, patients do not always respond in the same way to a given treatment. For example, half

of all patients currently do not respond to standard therapy for diffuse large B cell lymphoma; the reason is that there are various forms of the disease caused by aberrant molecular abnormalities, thus making the cancer more potent than usual. By using microarrays, physicians hope to be able to identify which patients carry this virulent type of cancer so that they can quickly initiate aggressive, intensive therapy on them (6).

Microarrays can also be applied by physicians to diagnose genetic disorders, which can successfully pinpoint the specific mutation on the gene locus responsible for the disease. By accurately characterizing the nature of the mutation, physicians will be able to prescribe treatments that can optimally benefit the patient. One such example where such a strategy could be implemented is in cystic fibrosis. Though it is known to be caused by mutation in both alleles of the CFTR gene, there are over six hundred possible mutations in the large CFTR gene. These mutations can range from the inability of a protein to fold properly to a protein that does not respond to second messengers. Since the disease severity is dependent upon the position and nature of the mutation on the CFTR gene, different therapeutic approaches for cystic fibrosis can be more effective when matched to the corresponding mutation. Current development is underway for a microarray chip which can quickly screen for all the known CFTR mutations in the patient's CFTR gene. Such a chip would facilitate the accuracy of diagnosis of patients by doctors (2).

Another exciting development in this field of personalized molecular medicine is using protein microarrays, based on the understanding that the dysfunction of protein interactions is the most direct, underlying cause of diseases as they oversee all biological processes and cellular fates.   The interaction amongst a network of proteins is

analogous to the interaction of multiple genes described in gene expression profiling

above, differing only in that proteins act downstream to genes. Gene mutations result in

aberrant protein expression, giving rise to abnormal phenotypes. These protein

microarrays work very similarly to DNA microarrays with the exception that protein

arrays use proteins instead of cDNA on their surfaces. These microarrays can be

designed to profile the state of already known signal pathways or defined networks

within cells; this is accomplished through labeled ligands that can recognize an activated

state of each protein (7).

Such clinical proteomics will be able to significantly aid physicians in the future

by guiding diagnoses and therapeutic decisions. Through simple procedures such as a

diagnostic biopsy or looking at body fluid samples, doctors can become capable of

detecting disease early by analyzing proteomic patterns in the samples. Then, this

proteomic information could be further used to select a therapeutic combination that will

target the protein network involved in the disease, optimize therapeutic efficacy, and

minimize toxicity and drug resistance (7).

In spite of all the potential that microarrays have to offer, some obstacles remain

that must be addressed to make this technology successful. A fundamental principle

driving microarrays is that both diseased and normal samples can be compared for

differences during the course of a disease, or in response to a treatment. The definition of

"normal," however, is ambiguous at best; when conducting expression analysis for

disease, one faces the dilemma of either comparing a patient's diseased tissue to that

same patient's healthy tissue, or to the healthy tissue of a "control" individual (8). With

the latter choice, there are concerns that gene expression profiles can vary due to the race, sex, age, and genetic makeup of the "control" individual in relation to the test subject. With the former choice, complications could arise from the proximity of the normal tissue that is used as the control. It is very possible that tissue adjacent to a diseased area could have already had an altered expression profile, even though it appears healthy. In addition to the variability of expression profiles in different tissues of the same organ, these heterogenous tissues are composed of early /late-stage cells and different cell types that each carry a characteristic pattern of gene expression (8). These obstacles, however, should be surmountable and are overshadowed by the overwhelming potential that microarrays present . For example, 3-dimensional modeling is already being employed to address the concerns of defining "normal." This technique can divide a specimen into transverse cross sections to be able to easily differentiate normal versus diseased tissue.

Microarrays represent an unprecedented departure from conventional technology in the field of functional genomics.   These chips can gather data at a significantly rapid rate and facilitate analysis through comparisons of genes or proteins that are represented on the microarray. The possibility of being able to conduct an expression analysis of the entire human genome in one simple test is a vast improvement over the belabored analysis of a few genes at a time using current, genomic approaches. With the variety of applications feasible in areas ranging from expression analysis to drug development to disease prognosis/diagnosis, microarrays offer a novel way to be able to revolutionize our understanding of the biological processes associated with the human genome.

REFERENCES

1. Schena, M., Heller R.A., *et al.* "Microarrays: biotechnology's discovery platform for functional genomics." <u>Tibtech</u> 16 (1998): 301-306.

2. Wallace, R.W. "DNA on a chip: serving up the genome for diagnostics and research." <u>Molecular Medicine Today</u> 3 (1997): 384-389.

3. Brown, P.O. and Botstein, D. "Exploring the new world of the genome with DNA microarrays." <u>Nature Genetics</u> 21 (1999): 33-37.

4. Lockhart, D.J. and Winzeler, E.A. "Genomics, gene expression and DNA arrays." <u>Nature</u> 405 (2000): 827-836.

5. Debouck C. and Goodfellow, P.N. "DNA microarrays in drug discovery and development." <u>Nature Genetics</u> 21 (1999): 48-50.

6. Friend, S.H. and Stoughton, R.B. "The Magic of Microarrays." <u>Scientific American</u> (2002): 44-53.

7. Liotta, L.A., Kohn, E.C., *et al.* "Clinical Proteomics." <u>JAMA</u> 286 (2001): 2211-2214.

8. King, H.C. and Sinha, A.A. "Gene Expression Profile Analysis by DNA Microarrays." <u>JAMA</u> 286 (2001): 2280-2288.