Microarray applications and challenges: a vast array of possibilities

Ahmed Fadiel #1 and Frederick Naftolin 2

1 The Center for Applied Genomics, Hospital for Sick Children, 555 University Ave., Toronto, Ontario, M5G1Z8, Canada
2 Yale University School of Medicine, OB/GYN Department, New Haven, CT, USA

# Corresponding author. E-mail: afadiel@bioinfo.sickkids.on.ca

Received: 2002-12-6; Reviewed: 2003-01-13; Revised: 2003-01-16; Accepted: 2003-01-25; Published: 2003-02-02

Keywords: Microarrays, High throughput, Molecular toxicology, Drug screening, Molecular medicine and Gene expression analysis

Abstract: In the past several years, a new technology known as microarray has attracted tremendous interest among biologists and biomedical researchers. The introduction of this technology into biomedical research has significantly affected the way questions about diseases and/or biological phenomena are addressed. This is because microarrays facilitate monitoring hundreds, even thousands of genes on a single chip giving scientists a better picture of simultaneous interactions among these genes. Such holistic views were hard to obtain using conventional molecular biological technologies in which one gene is investigated in one experiment, providing an incomplete picture of how genes function and interact. It is believed that obtaining such holistic views with microarrays will revolutionize and reshape biology; scientists are able to track down thousands of genes and their products in a given living organism and to understand how these genes function in an orchestrated manner. Here, we provide an overview of microarray technology, its technical aspects, general principles, and its applications in Bioscience and highlight its major biomedical applications. We also look at the challenges and problems facing this remarkable technology.

2. General introductory remarks

2.1 Arrays and microarrays: historical perspective

Next came the use of Southern blotting to screen filter lifts of cDNA libraries, where each bacterial clone represents a single cDNA species. Gridded libraries provided a means of arraying bacterial colonies, with a fixed position for each gene. The degree of miniaturization of the sample determines whether the array is a macroarray and microarray. Currently available “macroarrays” on nylon membranes allow analysis of several hundred genes at once; however, significant improvements in substrate materials, robotics, and signal detection have made possible miniaturization of arrays that have hundreds of thousands of oligonucleotides arrayed on a cm² chip. Microarrays make it possible to survey gene expression in a systematic comprehensive and efficient manner [4].

The use of microarrays to study gene expression profiles in biologic samples began in 1995 [5]. This breakthrough in microarray technology came about through two key innovations: First was the use of solid supports, such as glass or silicon chips that facilitated miniaturization with fluorescence-based detection that allows analysis of tens of thousands of genes in a few cm². The second major breakthrough was the development of methods for high-density oligonucleotide synthesis directly on the microarray chip. The development of microarray technology is dependent on both advances in the acquisition of genetic sequence information and the ability to study small quantities of nucleic acid [6].

Although microarrays can be used to study DNA, RNA or protein expression, the term generally refers to cDNA testing for RNA expression rather than to protein for which the term protein array is more common or for DNA for which the term genomic array is used. The term expression arrays is usually used to denote testing for RNA expression [7]. Although genomic and protein arrays are important applications of microarray technology, they are not the main subject of this manuscript.

2.2 Microarray fabrication and how arrays work

Despite the availability of expression arrays in many varieties, generally two categories are employed; customized cDNA microarrays that are composed of cDNA or oligonucleotides and commercially produced high-density arrays that contain synthesized oligonucleotides [8]. Whether they are developed by a company or non-commercially by a scientist arrays depend on the same basic principle: The ability of a single stranded sequences of nucleotides to stick to-
“hybridize” to its complementary sequence and form double-stranded DNA. Spotted arrays such as shown in [Figure 1] can compare RNA’s from two samples on a single chip. To do this they require a source of genes spotted on the chip. These are usually furnished as Expressed Sequence Tag’s (EST) clones or oligonucleotides. Briefly, microarray systems are usually comprised of cDNA probes formatted in a micro-scale on glass surfaces (chip) plus the instruments needed to handle samples (automated robotics) and to read the reporter molecules (done by scanners) and to analyze the data (using Bioinformatics tools). Commercially available high-density arrays provide analysis of the expression of a large number of genes (eg. 12,000 human genes on the chip); yet they can only analyze a single sample per chip. This means considerable cost and makes them unsuitable for large-scale experiments in most laboratories [4].

Microarray systems: The basic types of DNA microarrays include sequencing chips, expression chips and chips for comparative genomic hybridization. Technologies used in automated microarray production are photolithography, mechanical micro spotting and ink jets [Figure 2]. Generally, the applications of this technology are either identification of gene mutation or determination of levels of gene expression. Monitoring of gene expression or the occurrence of polymorphisms in genomic DNA is made through the hybridization of RNA- or DNA-derived samples on chips. The hybridization of labeled nucleic acid molecules to an array of complementary sequences immobilized on a solid substrate is the key unifying principle of all microarray experiments [9, 10].

In greater detail, a simple cDNA array experiment has five basic steps. First, the target cDNA is spotted or printed onto a solid substrate (glass or plastic surface). Then the sample RNA is isolated. Third, the cDNA is synthesized including labeling it for later detection. Fourth, the labeled probe cDNA is hybridized to target the cDNA on the substrate. When sample DNA or RNA is applied to the array any sequences in the sample that find a match will bind to a specific spot on the array. Finally the hybridization results are imaged and analyzed. Using specialized computer programs the amount of sample bound to each spot on the microarray is determined. A simple schematic drawing is shown in [Figure 2], in which the microarrays procedures are explained in a sequential order. The results are color-coded, so that the most active genes are colored red, and genes that are repressed are colored green. Alternatively, the experiment can be set up so that genes that are over represented in cancer cells are labeled red and genes that are missing in the tumor cells are labeled green. In summary, DNA microarrays use the same DNA probe detection method as in single gene studies but on a much larger scale that allows thousands of specific DNA-or RNA sequences to be tested/detected simultaneously.

Chip construction: Although several methods for building chips have been developed [11-13], two have especially prevailed. In the more commonly used method microarrays
are constructed by physically attaching DNA fragments such as library clones or polymerase chain reaction (PCR) products to a solid substrate [8, 14, 15]. By using a robotic arrayer and capillary printing tips at least 23,000 gene fragments can be printed on a microscope slide. In the other method arrays are constructed by synthesizing single-strand oligonucleotides in situ by using photolithographic techniques [8, 16]. The advantages of the former method include relatively low cost and substantial flexibility, while the advantage of the latter method is the higher density (>280,000/1.28X1.28- cm array) and elimination of the need to collect and store the cloned DNA or PCR products.

Customized cDNA microarrays are fabricated by selecting the genes to be printed on the array from a (public) databases/repositories or institutional sources. A high-throughput DNA preparation, usually done by robotic systems, performs the necessary tens of thousands of PCR reactions. Purified PCR products representing specific genes are spotted onto the chip (matrix) by a robot, which deposits a nanoliter of PCR product onto the matrix in a retrievable order. Nylon filter arrays have largely been replaced by glass-based arrays, typically microscope slides, which have the advantage of two-color fluorescence labeling with low inherent background fluorescence. DNA adherence to the slide is enhanced by treatment with polylysine or other cross-linking chemical coatings. Spotted samples are cross-linked to the matrix by ultraviolet irradiation and denatured by exposure to either heat or alkali [8].

2.3 Molecular biology and microarrays

In the mid-1970s molecular biology started a new era with their major focus on developing molecular cloning and DNA sequencing. Five years later, automated DNA sequencing took the world of biotechnology by storm with the dream of sequencing the entire human genome. This dream later seemed to be possible with the development of automated sequencing techniques and the invention of the polymerase chain reaction (PCR), which produced a surge in new experiments [17]. 2001 witnessed a major achievement through the declaration of the completion of the first draft of the human genome sequence. Science observers described this as the greatest scientific achievement since the discovery of the atom and the gene. It was even mentioned as the greatest accomplishment since the discovery of agriculture. Having the human genome gave a profound impetus to the strategies biological scientists can use to identify disease-associated genes. The Human Genome Project has created a massive amount of DNA sequence information and influenced many research and educational fields [18-23].

Besides the human genome, other prokaryote and eukaryotic complete genome sequences have also become available. This massive growth of information created a new challenge in the form of new gene identifications, determining the function of these genes, defining disease associations, and elucidating the correlation between genotype and phenotype [24]. To overcome these challenges a major aim for bioscience workers has been to study genetic variation in large sets of genes; “not one gene at a time”. There is hope that this will illuminate key single or clustered genes, expose important disease mechanisms, identify novel drug targets and also enable the prediction of the therapeutic responses to these drugs [17].

With the massive increase in molecular information and gene sequences, the tradition of analyzing one gene at a time in not able to give the comprehensive answers for many questions such as gene co-regulation and interactions. There is need for developing technology by which clusters of genes can be analyzed simultaneously.

New techniques and tools for conducting research have been developed to overcome these technical obstacles and allow full advantage of the enormous amounts of genomic information. This technology is the microarray, which allow simultaneous large-scale expression measurements on great numbers of genes using minute samples. Despite the current widespread use of microarrays, there is no single terminology on which science community agrees. Descriptive terms that have been used generally describe the technology itself. These include but are not limited to microarray, biochip, DNA chip, gene array, gene chips and more recently genomic chips (considered by some to include protein chips that mark protein expression using labeled antibodies).

DNA microarrays, which are also, called DNA arrays or gene chips, are an example microarrays that use genome sequence information to analyze the structure and function of tens of thousands of genes at a time. Thus, biotechnology laboratories and pharmaceutical companies can identify the molecular targets with which drugs actually interact. Since it can also identify individuals with similar biological response patterns, microarray analysis can assist drug companies in choosing the most appropriate candidates for participating in clinical trials of new drugs. In an extension of these methods, in the future arrays will help medical professionals select the most effective drugs for individual patients.

3. Applications Of Microarray Technology

As previously mentioned, gene chips facilitated a more comprehensive and inclusive experimental approach in which alterations in the state of entire genomes simultaneously can be assayed during health and disease or in response to a variety of stimuli [Figure 3]. This ability to profile changes in gene-expression levels under different conditions makes microarrays the method of choice in many fields. Some of these fields are disease fingerprinting, drug targeting and evaluation, toxicity assessment, signal transduction research and developmental cancer treatment [25, 26]. All of these potential uses increase the demand for microarrays in both academic and industrial settings.

3.1 Medical and clinical diagnosis

The ability of microarrays to provide genome-wide profiles of gene expression increases potential clinical applica-
Microarray applications and challenges

3.2 Pharmacological industry and microarrays

Drug discovery, targeting and monitoring with microarrays

There are three major tasks with which the pharmaceutical industry deals on a regular basis: (1) to discover a drug for an already defined target, (2) to assess drug toxicity, and (3) to monitor drug safety and effectiveness. As well, determining the potential problems of a compound early in the drug discovery process will save time and money by focusing resources on compounds that are more likely to succeed. Additionally, the success of drug screening depends both on the speed or turnaround-time of evaluation and on the reliability of results [32, 33]. More to the point, for a long time one of the most pressing issues facing the pharmaceutical and biotechnology industry was the high dropout rate of lead drug candidates. Late-stage dropouts are both financially costly and undermine confidence in the company [34]. Relief was found with the increase in the knowledge of total genomic sequences of many prokaryotes and eukaryotes. These represent gold mines for drug targeting and discovery. Accordingly, there was an urgent need to develop high throughput monitoring technologies that are meant to identify targets and provide lead candidate optimization. Accordingly, there was a call for developing a technology that can help in basic research, drug discovery and evaluation [35].

Gene expression microarray allows the simultaneous monitoring of the expression levels of thousands of genes [8, 15, 16, 36]. This is an advantage that makes microarrays suitable for measuring the expression patterns of thousands of genes in parallel and generating clues to gene function that can help to identify appropriate targets for therapeutic intervention. Microarray also can be used to monitor changes in gene expression in response to drug treatments [35, 37]. This makes gene chips the current frontrunner technology in the drug industry. Moreover, this also provided prospective markets for this infant technology.

With their ability to integrate genomics, pharmacology and molecular medicine, gene chips promise to revolutionize these fields as they reveal clusters of genes and gene regulation events involved in disease progression and to pinpoint potential drug targets and diagnostics [38, 39]. By comparing the ways in which genes are expressed in a normal and a diseased heart, for example, it is possible to identify the genes and hence the associated proteins that are part of the disease process. Researchers could then use that information to synthesize drugs that interact with proteins, thus reducing disease’s effect on the body. Moreover, since that
many drug targets are components of complex signaling pathways, and in order to understand the true biological consequences of modulating these targets it is necessary to understand the biology of the system in great detail. Since activation of signaling pathways leads to mRNA expression, microarray technology can be used to provide a detailed quantitative assessment of the consequences of this activation, providing a new biological perspective on “well-established” cellular systems [40]. Because pharmacogenomics is the hybridization of functional genomics and molecular pharmacology it also aims to find correlation between therapeutic responses to drugs and the genetic profiles of patients as described above. These efforts will allow answers to questions about the effects of certain drugs on certain individuals, or ethnic groups, and identify appropriate targets for therapeutic intervention [25, 38].

Determining the potential toxicity of compounds early in the drug discovery process can be extremely beneficial in terms of both time and money saved. Moreover, to accurately evaluate the large number of compounds being produced, toxicology assays must have both high fidelity and high-throughput capabilities. Assays must be performed using limited amounts of compound though. Conventionally with traditional techniques these were very hard tasks [41-51]. Currently, using microarrays to fulfill these goals and obtain detailed information concerning the molecular mechanisms behind toxic effects is becoming routine procedures in pharmaceutical companies replacing many conventional techniques. Besides being able to provide information on thousands of gene in one experiment which saves time it is rapid, cost effective, and relatively low-priced [41-53].

3.3 Microarrays applications in toxicogenomics
Toxicogenomics is the hybridization of functional genomics and molecular toxicology. The goal of toxicogenomics is to find correlation between toxic responses to toxicants and changes in the genetic profiles of the objects exposed to such toxicants (see above). Currently, microarray technology is gaining popularity in predictive toxicology [39]. The role of gene chips is to allow monitoring changes in gene expression, where change in mRNA or protein levels enables investigators to reject or embrace hypotheses about pharmacology or toxicology [32]. This facilitates testing drug toxicity, safety and effectiveness which are very critical where, determining the potential problems of a compounds early can save time and money by focusing on compounds that are more likely to succeed. Moreover, the use of animals for in such studies is minimized because of the percussion and conservative nature of the testing [41, 43, 47, 52]. On the other hand, the possibility of utilizing microarrays in exploring genes from other organisms will allow practical studies of questions about basic cellular functions, including cell cycle and DNA repair mechanisms, which are the major issues in cancer and cytotoxicity research. Since major signaling pathways and cellular processes involved in cellular response to cytotoxic agents are conserved between organisms, as in yeast and mammals, these simple eukaryotic systems could be excellent models for testing sensitivity to e.g. anti-tumor drugs [54].

Utilizing microarrays helped significantly to overcome many long-standing challenges when drug testing was required for organs with complex expression patterns or in certain illnesses. As one of these organs is the human brain, for which measurement of gene expression with microarrays has potential applications in the study of many diseases especially cognitive disorders for which animal models are typically not available [55].

3.3 Microarray based oncology
Cancer is a genetic disease and as such, our understanding of the pathobiology of tumors derives from analyses of the genes whose mutations are responsible for those tumors. The cancer phenotype, however, likely reflects the changes in the expression patterns of hundreds or even thousands of genes that survived the primary mutation of an oncogene or a tumor suppressor gene. Recently developed functional genomic approaches, such as DNA microarrays have enabled researchers to determine the expression level of every gene in a given cell or cancer population, which represents that cell population’s entire transcriptome [56-58].

The application of microarray and related technologies to identify specific targets of defined genes that have clearly been implicated in cancer progression requires different experimental approaches than are conventionally used. The objective is to define changes in transcriptional profiles that occur in response to modulating the expression level of the gene to be studied. The resulting altered expression profile can then be viewed as a blueprint by which that gene affects its cellular function. In this approach the investigator has greater control over the question being asked as the experimental variables are artificially controlled. Therefore multiple repetitions of the experiment are not required and in general the experiment carried out in duplicate is sufficient to generate reliable data. Despite the fact that many techniques were used in gene profiling experiments, such as differential display [59-62] and serial analysis of gene expression [63-68].

Currently cancer research coupled with diagnostic DNA microarrays is playing a dominant role compared to the other developing technologies since they are relatively easy to make and use and are applicable to numerous scientific inquiries. In addition, microarrays’ ability to simultaneously analyze several thousands of genes in samples from sick or healthy tissues will provide information that is expected to improve our understanding of the complex molecular interaction networks of healthy cells and tissues. Such characterization of the molecular mechanisms involved in pathology should result in the identification of new therapeutic targets and the development of new medicines. The genetic profiles thus obtained should also permit the definition of new pathologic subclasses not recognizable by traditional clinical factors, as well as new markers for susceptibility to cer-
Microarray applications and challenges

Microarrays' ability to provide a systematic method to identify key markers for prognosis and treatment response by profiling thousands of genes expressed in a single cancer is significant when dealing with certain cancer types. Two of these are breast cancer, which is the most frequent and deadly cancer in women and prostate cancer in men [57, 71]. The great heterogeneity of breast cancer makes prognosis and response to current treatments highly variable and difficult to predict. The use of expression profiling with cDNA array techniques in mammary tumor cell lines and breast tumors will help classify controversial tumors and provide new prognostic tools and potential therapeutic targets. They will also boost our knowledge of the molecular events responsible for the development and progression of cancer [39, 70, 72]. Moreover, this will also allow us to define specific molecular pathways of tumoral progression and to define markers of prognostic and diagnostic relevance, which will reflect significantly on cancer diagnostics and consequently its management [58, 73, 74].

4. Challenges

Microarray technology was initially limited to biotechnology companies and high-budget research laboratories. Gene-chip companies have developed marketing arrangements and collaborations to bring microarray chips to academic research laboratories as well as commercial users. However, microarrays continue to have limitations in addition to their technical difficulty, specificity, and reliability: (4.1) Microarrays are victims of their own success since the large data sets generated by the chips add new statistical and Informatics-related challenges and complexity. (4.2) There is still no standard protocol for microarray data analysis. (4.3) There is no uniform system for generally managing and provide a disbursement point for microarray data and for the products of various data-mining techniques and their applications. Today, while microarray experiments will undoubtedly push forward our knowledge about expression and regulation of genes they still raise many technical and informatics challenges. (4.4) Novel statistical methods must be evolved for the analysis of these large and diverse datasets.

4.1. Microarray Data Handling and Management

Microarray technology has resulted in the generation of large complex data sets. This situation shifted the bottleneck in biological investigation from data generation to data analysis [26, 75-79]. This data-handling and analysis has become one of the major bottlenecks in the utilization of array technology. The amounts of data obtained from each experiment can require powerful computing techniques for identifying clusters of genes and interpreting overall patterns of expression. At present it seems that the greatest challenges in microarray research are not the arrays itself but the way by which resulting data matrices are handled and analyzed. Without the ability to handle the data reliably and expeditiously the development of knowledge about the underlying biological processes, etc. will not be practical [80-83].

4.2 Lack of a standardized protocol for data handling

Once the hybridized chip is scanned, the data flow encounters the following impediments: Data are collected and saved as both image-and text files. It is of critical importance that databases and tracking files are maintained by precise registration of the spot configuration of all genes on the chip. Then the saved files can safely be imported to software programs for performance of image analysis and statistical analysis functions without concern over possible overlap, etc.

The data are mined for induced or repressed genes, patterns of gene expression, and temporal relationships of expression under different experimental conditions. A significant challenge exists in making sense of the vast quantity of data generated by microarray experiments. There is no single tool that meets all of the needs of the microarray researcher. Mixtures of software programs presently are used to perform a multitude of tasks, including data tracking, image analysis, database storage, data queries, statistical analysis, multidimensional visualization, and interaction with public databases on the Internet [84-86]. Basic spreadsheet programs have been adapted to answer questions regarding the magnitude of change in gene expression. However, limitations often arise as a result of mismatched programs and inadequate memory capacity for managing the enormous data sets. It is only recently that more sophisticated and comprehensive analytical tools for cluster analysis, self-organizing maps, and principle component analysis have been applied to biologic data to extract higher-order relationships embedded in expression patterns [4, 84].

4.3 Informatics Challenges

One of the challenges facing current microarray technology is the absence of many gene annotations and the inability to find with ease literature references to relevant genes. This extends to difficulties in linking observations with the actions of the products of the genes being studied, including the possibility of gene clusters. Future challenges for microarray research will therefore include development of databases and algorithms to manage and analyze vast scale genomic datasets [87]. Some pertinent examples of available databases are shown in [Table 1].

Finally, the lack of a standard format for microarray data interferes with the creation of such a resource [87, 88]. This problem was recognized by the scientists at the European Bioinformatics Institute when they proposed that standards should be based on XML, a computer markup language that combines data and formatting in a single file for distribution over the World-Wide-Web [89].

4.4 Statistical challenges: array technology is not old wine in new bottles and the same must be true for the statistical
### Microarray applications and challenges

Table 1. Major Microarrays resources available on web.

<table>
<thead>
<tr>
<th>Project name</th>
<th>URL address</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Bioinformatics Institute</td>
<td><a href="http://www.ebi.ac.uk/arrayexpress/">http://www.ebi.ac.uk/arrayexpress/</a></td>
<td>Arrayexpress</td>
</tr>
<tr>
<td>NCI-LECB</td>
<td><a href="http://www.lecb.ncifcrf.gov/MAExplorer/">http://www.lecb.ncifcrf.gov/MAExplorer/</a></td>
<td>MAExplorer</td>
</tr>
<tr>
<td>Stanford Microarray Database</td>
<td><a href="http://daisy.stanford.edu/MicroArray/SMD/">http://daisy.stanford.edu/MicroArray/SMD/</a></td>
<td>SMD</td>
</tr>
<tr>
<td>National Human Genome Research Institute</td>
<td><a href="http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/">http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/</a></td>
<td>ArrayDB MAPS</td>
</tr>
<tr>
<td>Manchester Bioinformatics</td>
<td><a href="http://www.bioinf.man.ac.uk/microarray/maxd/maxdView/index.html">http://www.bioinf.man.ac.uk/microarray/maxd/maxdView/index.html</a></td>
<td>Maxd</td>
</tr>
<tr>
<td>Yale Microarray Database</td>
<td><a href="http://info.med.yale.edu/microarray/">http://info.med.yale.edu/microarray/</a></td>
<td>YMD</td>
</tr>
<tr>
<td>San Diego Supercomputer Center (SDSC)</td>
<td><a href="https://array.sdsc.edu/">https://array.sdsc.edu/</a></td>
<td>2HAPI</td>
</tr>
<tr>
<td>University of Pennsylvania</td>
<td><a href="http://www.chil.upenn.edu/RAD2/">http://www.chil.upenn.edu/RAD2/</a></td>
<td>RAD</td>
</tr>
<tr>
<td>George Church Lab’s (Harvard)</td>
<td><a href="http://twod.med.harvard.edu/ExpressDB/">http://twod.med.harvard.edu/ExpressDB/</a></td>
<td>ExpressDB</td>
</tr>
<tr>
<td>Pevsner Lab’s (Kennedy Krieger Institute)</td>
<td><a href="http://pevsnlab.kennedykrieger.org/microarraydata/">http://pevsnlab.kennedykrieger.org/microarraydata/</a></td>
<td></td>
</tr>
<tr>
<td>Young lab (White head institute-MIT)</td>
<td><a href="http://young39.wi.mit.edu/chipdb_public/">http://young39.wi.mit.edu/chipdb_public/</a></td>
<td>Dragon chipDB</td>
</tr>
<tr>
<td>Another Microarray Database</td>
<td><a href="http://www.microarrays.org/software.html">http://www.microarrays.org/software.html</a></td>
<td>Microarraysorg</td>
</tr>
</tbody>
</table>

methods that deal with array data.

Despite the previously mentioned problems, obtaining data with gene chips has already become routine in many laboratories. However, large scale array data meets with the resistance or inappropriateness of conventional methods of data reduction and analysis: interpreting the data is a major difficulty.

Firstly, there are serious difficulties in registering and solving microarray images. New types of software and statistical techniques are required for this [84, 90, 91]. The US government represented by the National Institute of Health (NIH) has made programs to support such developments (http://research.nhgri.nih.gov/microarray/index.html and http://grants1.nih.gov/grants/guide/rfa-files/RFA-NS-02-001.html).

Next, the usual paradigm in bio-statistics, few variables with many cases, is turned topsy-turvy in microarray work. In the latter the question is “which are the variables that are significant?” Arrays can examine thousands of genes while virtually nothing is known about their functions and expressions. The problem of multidimensional applications of the data can be solved by building an understanding of what is (are) the function(s) of the genes that are the subject of analysis and is there any discernable logic to their expression. This problem is likely to be resolved by the development of matrices from biological and computational studies; it is a byproduct of the solutions. At present, it is more critical for microarrays to be sufficiently uniform that the data from one study can be displayed and solved along with other studies. In general it can be said that genes that are co-expressed are highly likely to participate in similar physiological programs in physical processes. This understanding will facilitate the design of novel, more accurate and intuitive statistical methods. This will not be solved by computation alone; the statistical problem of going backwards from thousands or tens of thousands of measurements to fundamental biological processes is not a soluble mathematical problem. But, known interactions can be mapped onto the data that is being generated from these very large-scale experiments.

On the other hand, because microarray experiments result in such large amounts of data, false-positive results are likely to be obtained, where, analyzing multiple independent experiments may eliminate spurious results. As a possible solution, a transition was made from experiments involving a small number of conditions (with an emphasis on the specific genes induced or repressed) to experiments involving hundreds of conditions. In these experiments patterns of global gene expressions are used to classify disease specimens and discover gene functions and drug targets [92]. Currently, many groups are developing software that will help to interpret microarray data using non-conventional statistical techniques.

The present methods of managing data are not sufficiently precise to allow proper statistical analysis. The use of ‘fold difference’ (the ratio of the expression in cells of interest versus the control cells) as a quantitative measure of the differential expression is an example of this problem. Genes expressed at low levels require higher fold differences in order to rise above the level of the noise in the systems and furnish proper signal to noise ratio calculations [93, 94].

The present method of resolving this problem is performing replicates to minimize the effect of the noise in the system. But, this does not deal with the fundamental problem in the signal to noise ratio. There are other methods for assigning confidence for differentially expressed genes [95]. But the noise in the data can still confound analyses and rank data are often more robust than absolute measured values. All of this is due to the relatively high noise to signal ratio in microarray techniques at this time. Either the systems must be made more precise or new statistical methods must be developed to correct for the noise without affecting the impact of the signal [96, 97].

---

www.iabs.us  Int Arch Biosci 2003: 1111-1121  1117  Fadiel A, Naftolin F
There is also a problem stemming from gaps in the data flow from microarray assays. Methods have emerged for imputing incomplete data, which may mislead the outcome of the analysis [98].

Finally, two problems generic remain major concerns for (statistical) analysis of array data, just as they are problems for any gene sampling exercise. First is aneuploidy (the number of copies of a gene in the effective genome), which has been shown to effect the measured expression level of a gene. This in turn either confound the analysis or inhibit insight into mechanisms of abnormal biology [99-102]. Second, is the occurrences of false positives that rises sharply when thousands of genes in a microarray data set are evaluated simultaneously by fold changes and significance tests [103]. Although beyond the scope of this review, these are serious, organic problems for gene analysis and a system must be devised to flag the data they produce and cleanse it from the data sets prior to final analysis.

5. Concluding remarks

Microarray technology is becoming a corner stone in the development and integration of knowledge obtained from molecular targets [9, 26, 70]. There are many microarray techniques with clinical applications. By facilitating exploration of the relationship of genes to disease, and then to the analysis of the information great inroads will be made to the treatment of patients as well as the accumulation for its own sake. Arrays have an important dimension to the breath and depth of knowledge that is obtained. The future of microarrays seems even more promising to the life sciences than PCR technology [10, 17]. Its potentials are only limited by the practical constraints that we have mentioned briefly in this review. This technology will allow the application of new insights, some of which are self generated. At present the use of microarrays studying variations at the single nucleotide levels (SNPs) for individuals for gene profiling has allowed the characterization of profiles that are consistent with the development or presence of cancer. That they will produce the ability to rule out the likelihood of cancer remains to be proven [69, 71, 72, 104]. Coupled with advanced Informatics, microarray diagnostics will furnish individualized investigation of the physiology and disease patterns of individuals and may be able to dispense the need for large, costly and often difficult to interpret clinical trials [20, 81]. They may lead to true individualized prevention and treatment of diseases [88, 104, 105]. The integration of genes and environmental factors now possible is thought to be critical not only for cancer, but for many other diseases, such as diabetes, heart disease, asthma, and some neurological disorders [93, 106-110]. In addition, this will radically improve the identification of both genetic and molecular causes of susceptibility to certain diseases [27, 30, 88, 111]. Array technology has already revolutionized fields such as pharmacogenomics. There was a clear progress in finding drug targets and hence drug discovery, drug development and safety and molecular diagnostics were very clear [29, 32, 35, 53, 105].

Despite these phenomenal accomplishments and further potential there was a major drawback to defeat; microarrays are entirely dependent on the state of knowledge of the genome under investigation. But this is fading as a principal problem due to the solution of the human genome and the speed at which various other genomes have been or are in the process of being sequenced. Another criticism that has yet to be resolved is the concern that the expression levels achieved in these artificial systems are not physiologically relevant. This is undeniably of concern and the onus is on investigators to devise additional experiments to confirm that genes identified in this manner represent physiologically relevant targets [26, 112].

We have also mentioned several methodological challenges to the practicality of microarrays; but there seems little concern that these problems will yield in the foreseeable future.

Looking at the larger picture it is clear that analysis of gene expression is only in its infancy. Even the rather obvious approaches, such as cluster analysis and finding differentially expressed genes, have been used only rather crudely. It is expected that gene-expression data analysis methods will develop similarly as sequence analysis methods have developed over the past decades. The amounts of gene expression data will continue growing and the data will become more systematic. Currently gene expression profiling is similar to gene sequencing before the era of genome sequencing: the measurements are carried out to attack particular questions or sometimes just to demonstrate the concept. Establishing a public repository for gene expression data would facilitate systematic construction of gene expression matrices for various organisms [15, 89]. Like genome sequencing, the systematic gene expression profile is not an end in itself. It is a long way from having detailed gene expression profiles to the real and practically usable understanding of underlying cellular processes. Bioinformatics methods and tools will be needed to cope with the huge amounts of data, but they will not bring any deep understanding by themselves. On the other hand, the traditional ‘gene by gene’ methods will not be sufficient to understand increasingly complex gene regulatory networks consisting of thousands or tens of thousands of genes. Microarray technologies are improving to the point when an entire genome will be monitored with very high resolution. With such potentials, researchers can explore thousands of genes faster and more cheaply [25, 26]. With all these arguments a key question remains to be answered, what do we need to do? The most pressing need is for new methods of bioinformatics, especially data accession, reduction and analysis, which must be developed to handle the flood of new information. Secondly, as far as we deal with biological processes through the genes that are involved in these processes we must use all means available to understand these processes and as a result to be able to interpret the data obtained by microarrays.

Microarray applications and challenges

www.iabs.us Int Arch Biosci 2003: 1111-1121 1118 Fadiel A, Naftolin F
We believe that microarrays are on their way to revolutionize Bioscience and to allow us to understand how life really works. We have opened a new and larger window on the secrets of life, but this has only shown us how much more there is to learn.

Reference
Microarray applications and challenges


