

Microarrays — ‘Chipping’ in Genomics!

Biju Joseph, Anupama Shrinivasan and Govindasamy Kumaramanickavel*

Department of Genetics and Molecular Biology, Vision Research Foundation, Sankara Nethralaya, Chennai 600 006, India

Received 30 August 2001; accepted 20 November 2001

DNA microarrays or gene chips are an assemblage of assorted short sequences of DNA or polypeptides embedded onto a solid medium such as glass or plastic slides/silicon wafers or nylon membranes. A state of the art technology in the elucidation of molecular basis of gene expressions, this versatile tool finds many applications ranging from assessing the burgeoning sequential information generated by the human genome project, to the grading of tumours and evaluating clinical prognosis. The present review elaborates the historical background of DNA microarrays, their construction as well as discusses with relevant examples, their immense potential in diagnostic and therapeutic research and the drawbacks encountered in their application.

Keywords:cancer biology, DNA microarrays, DNA chips

Introduction

The rapid progression of the Human Genome Project has enabled the creation of a vast repository of genetic information. It has since been an ongoing quest for researchers worldwide to translate this formidable resource into applications from which a better insight about the molecular basis of gene expression can be gleaned. In this regard, the invention of DNA microarrays can be considered as a benchmark in elucidation of the complex biochemical and molecular functioning of genes at the cellular level and their contribution to the onset of diseases.

DNA microarrays or Gene chips as they are popularly known are literally speaking short sequences of DNA or peptide nucleic acids, embedded onto a solid support such as glass or plastic slides, silicon wafers or nylon membranes. A typical microarray analysis involves the exposure of an ‘Immobile Phase’ that could be either PCR amplified genomic sequences, cDNAs, or oligonucleotides concentrated within a solid background, to a ‘mobile phase’ of fluorescently labelled DNA probe. The resultant binding of complementary DNA sequences is visualized as a ‘signal’ that is then counted for as an appraisal of gene expression. A single microarray unit allows for the surveillance of expression among thousands of genes from a single tissue specimen or that of a single gene in several tissues.

The potential of such technology is tremendous as it allows for the simultaneous study of several probable causative genes in disease and thereby obtains a more specific correlation between disease prognosis and gene expression. Microarrays can be used to study multiple molecular events in normal tissues such as those of embryonic development. Additionally, the assessment of the prognostic molecular responses to different therapies may be applied in the designing of newer drugs and treatment strategies (www.boland-pc.ces.clemson.edu/biochip/microarr.htm). Table 1 lists some of the fields wherein microarrays have been employed for analytical purposes successfully.

The present review explores the background of DNA microarrays, their fabrication as well as application in diagnostic and therapeutic research. It also discusses some relevant examples wherein the employment of this nascent technology has proved to be valuable to the investigation as well as highlights some of the pitfalls often faced in the appliance of this nascent technology.

History

The concept of DNA microarrays was first proposed in 1991 when the decoding of the human genome had just begun. In view of the vast amount of data predicted to be generated from the blueprint of the human genome, an acute necessity was felt for an application that could analyze several genes simultaneously and provide a clear picture of the molecular processes underway in a short span of time.

*Author for correspondence:

Tel: (44) 8271616; Fax: (44) 8254180

E-mail: drgkm@sankaranethralaya.org

Table 1—List of disciplines in biology and medicine applying microarray technology

Discipline	Field
Cancer	Cancer biology Cancer therapeutics
Developmental Biology	Tissue differentiation Gene mapping
Human Genome Project	SNP maps Genome annotation Cardiology Complex gene interactions (Disease related)
Medicine	DNA diagnostics Immunology Microbiology Neurodegenerative diseases Tissue typing Drug discovery Drug screening
Pharmaceutical industries	Toxicology

The idea was to merge the speed and accuracy of the automation technology with the analytical prowess of molecular biology.

In a path-breaking paper published in *Science*, Stephen Fodor and his colleagues at Affymax Research Institute, Duke University, Durham, North Carolina and Stanford University, Stanford, CA, reported on a process that used "solid phase chemistry, photolabile protecting groups and photolithography to yield a highly diverse set of chemical products" (Fodor *et al*, 1991). Employing the principles outlined in this pioneering paper on microarray technology, investigators at Stanford developed what is today traditionally known as a DNA microarray (<http://www.gene-chips.com>) in which probes (nucleic acid of known sequences) were immobilized on a solid surface such as glass and then exposed to a set of target DNA (free nucleic acid sample) separately and in a mixture (Ekins & Chu, 1999; Brown & Botstein, 1999). Capitalizing on the convenience of this newfound technology, Affymax later introduced the use of multiple protein blots on a silica surface for the purpose of drug screening.

In computer technology, the *silicon chip* is identified as an ideal tool for storing large amount of data for a fast paced macro- or micro-analysis. Employing the same principles in techniques for life sciences, investigators at Affymax Research Institute are credited with designing of a microarray unit known as the *DNA chip*, in which oligonucleotide sequences were impinged on a solid matrix. DNA-chip manufacturers are identifying newer mediums for use in the hybridization formats. Although most DNA-chip companies use a solid-phase technique in

their microassays, novel stratagems of using variations of both solid-phase and solution-based hybridization are being pursued actively.

One of the primary experiments carried out using microarray technology involved the delineation of the genes in *Saccharomyces cerevisiae* responsible for the metabolic shift from anaerobic to aerobic respiration (DeRisi *et al*, 1997). For this, yeast cells were cultured separately in conditions favouring either forms of respiration. The total RNA from the cells in these two cultures were isolated and then reverse transcribed into cDNA using a fluorescent dye which was green for anaerobically respiring cells and red for the aerobically respiring cells. Thereafter, a DNA microarray of all the 6400 genes in yeast was exposed to a cDNA fluorescent probe from the yeast cells. Following hybridization with the probe, the fluorescent signals generated by the complementary pairing of the genomic sequences with the probe were detected using a scanning laser microscope and the intensity of the signal analyzed by a computer for evaluating the extent of gene expression. Results of the study indicated that mRNA levels of mitochondrial protein important to aerobic respiration increased and those of the ribosomal proteins repressed. A total of seven genes showed a strong induction of expression when glucose levels were totally depleted. In general, the decrease of glucose as a source of carbon was accompanied by a twofold increase in gene expression in over 710 genes and twofold decrease in about 1030 genes and roughly about 1/4 of the yeast genome did not participate in the respiratory shift.

Microarray Fabrication

The construction of a microarray unit essentially involves the spotting or imprinting of the target DNA on a substrate, which could be glass, plastic, or a silicon wafer or even nylon membrane. There are generally three methods followed, which are discussed as under (www.css.orst.edu/Courses/css630/chip.htm; www.artsci.wustl.edu/~jstader/bloch.html; www.cs.washington.edu/homes/jbuhler/research/array/).

In situ Imprinting of "Probe DNA"

Oligonucleotide sequences are constructed based on their complementarity with the sequences of the genes of interest. With this information, computer algorithms are used to design photolithographic masks or receptacle sites for the arrays on a solid medium using a light-directed chemical synthesis process. This method combines solid-phase chemical

synthesis with photolithographic fabrication techniques employed in the semiconductor industry. Thereafter, using several chemical reactions to bond the oligonucleotide to the substrate a high-density array of oligonucleotides, with each probe in a predefined position in the array is imprinted onto a large glass wafer that is further diced before use. Individual probe arrays are then packaged in injection-molded plastic cartridges that protect them from the environment and also act as chambers for hybridization.

Piezoelectric Printing of Target DNA

An alternative approach involves the deposition of gene probes onto the chip substrate using a tiny droplet sprayer that resembles an ink-jet printer. A direct precise anchoring of nucleotides to a coated surface is then made using standard chemistry.

Spotted Array Printing

The spotted array-making technology involves the precise deposition of nanolitres or less of a concentrated DNA solution as "spots" on a substrate by a gridding robot. A latest development in this technology is the Pin-and-Ring mode of preparing spotted arrays that relies on the use of surface tension forces, for holding and transferring solutions whilst rapidly and uniformly placing precise volumes of fluids onto the surface of the substrate.

Microarray Analysis

A simple Microarray experiment involves the following steps (Fig. 1):

- The *Probe* which could be single, stranded cDNA or PCR amplified genomic sequences is imprinted onto a glass/plastic/silicon/nylon membrane substrate using either *in situ* or piezoelectric or spotted array printing methods.
- The probe is then exposed to the *Target* molecules, that is single, stranded cDNA or mRNA labelled with a fluorophore such as *fluorescein* or *Cy3* and hybridization of complementary sequences is carried out under highly stringent conditions.

Using a scanning electron microscope, emission spectra of the excited fluorescent dye tagged hybrids are detected and then measured. The consequent data is then imported into the software that generates a composite picture of colors. The data is basically viewed as a change in the fluorescent readings at a specific point on the slide (gene) as compared to a

reference sample. There have been several improvisations introduced presently in the detection of fluorescence during microarray analysis. The addition of confocal imaging technology permits a three dimensional detection of the fluorescence with a significant reduction in the background noise and a greatly improved sensitivity and accuracy in resolution. A novel method introduced recently uses Gold nanoparticles instead of fluorochromes for labeling the target molecules (Cheung *et al*, 1999; Duggan *et al*, 1999; <http://cmgm.stanford.edu/pbrown/mguide/index.html>).

Data Exploration

Analysis of data generated from the reading of a microarray involves the creation of a 'cluster algorithm' in which genes that show the same expression patterns or in other words co-regulated across the dataset are grouped together on the implication that they share common functions (Eisen *et al*, 1998).

Microarray Databases

Only few full-featured, relational gene expression databases have been developed due to lack of a standard format for consolidating microarray data. Recently, however, the European Bioinformatics Institute, Heidelberg, has proposed defining a standard based upon XML, a computer markup language that combines data and formatting in a single file for distribution over the World-Wide Web (<http://www.ebi.ac.uk/arrayexpress/>). Some of the other datasets available are <http://genomewww4.stanford.edu/MicroArray/MDEV/index.html> and <http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/dbase.html>.

Beaucage (2001) provides an excellent discussion on the different interfacial chemical reactions required to ensure the imprinting of oligonucleotides onto the surface of microarrays as well as the different modes of detection of fluorescence. Strategies for increasing the density of array as well as improvisations to study the kinetics and specificity of hybridization events such as the use of light-scattering techniques, molecular beacons, surface plasmon resonance, attenuated total internal reflection-FTIR, and the evanescent field excitation of fluorescence from surface-bound fluorophores have also been suggested.

Applications of Microarray Technology

The Microarray research is in a constant state of flux with innovations that strive to enable faster and more accurate analysis. The benefits of this new age

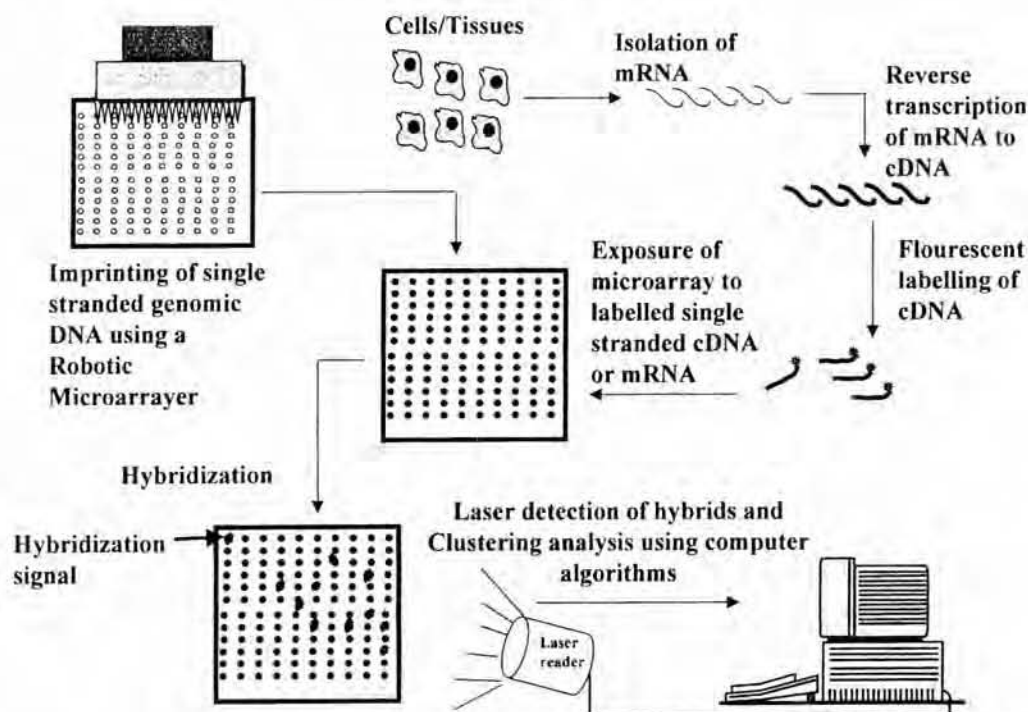


Fig. 1—A microarray experiment

technology are tremendous and there have been several accounts wherein the creative use of microarrays has provided an insightful picture of gene expression. An overview of some of the possible applications of this dexterous analytical tool is presented as follows:

The Human Genome Project

An important upshot of the ongoing sequencing of the human genome has been the identification of all coding regions. Such a meaningful interpretation of genomic sequences is also known as *annotation*, and it has enabled the compilation of an exhaustive catalogue of information on several genes, the transcripts they generate and their translated products. Traditionally annotation of the genome was accomplished using conclusions derived vis-à-vis non-experimental bioinformatic approaches like *ab initio* gene predictions, homology studies and motif analysis or comparative genomic approaches. The limitations of these methods lie principally in the time involved as well as the varying results obtained. While bioinformatic analyses provide a highly speculative account of number of genes, comparative studies between human and mouse genomes although identify 'potential candidate' genes cannot ascertain their activity status. In fact, Guigo *et al* (2000), concluded that despite more than 17 years of research

effort, the computational precise annotation of every human gene was a very distant goal. In this regard, the introduction of microarray technology could be an effective method of predicting exons as well as defining gene boundaries.

Chromosome 22 was the first human chromosome to be completely sequenced and computationally annotated. In an attempt to compare the traditional methods of computational annotation with microarray detection of gene expression, Shoemaker *et al* (2001) designed an ink jet array of overlapping oligos, called a *tiling array*, to monitor the activity of 8183 exonic regions annotated to the chromosome region 22q. For this they isolated mRNAs from human cell lines and normal and diseased tissue samples and fluorescently labeled them with two different dyes. Labelled targets were hybridized less than 69 pairs of experimental conditions with individual chromosome 22 exon arrays. Following detection, they developed a gene identification algorithm that compared log expression ratios with intensity. From this they were able to identify over 572 groups of exons (termed as expression verified genes or EVG) that shared a similar expression behavior across the conditional pairs as opposed to the 545 genes computationally annotated by Dunham *et al* (1999) in a previous study. Interestingly, from the 325 *ab initio* genes anticipated by Genscan, Dunham *et al* had predicted

only 100 of the transcripts to be representatives of true genes. However, the Shoemaker study found experimental support for over 185 genes. Additionally, they were also able to detect in a few cases that a single gene had been represented by more than one EVG, thereby suggesting the presence of an alternative splicing mechanism. This shows that the use of microarray technology could be an ingenious method of accelerating gene discovery.

One of the most stable and common DNA sequence variations across the human genome is the Single Nucleotide Polymorphisms (SNPs). An important utility of SNPs is that they can be used as markers in the identification of disease genes, wide risk association studies or also as monitors for individual differences to drug responses. Therefore, a complete map of SNPs would be an indispensable tool for genetic studies. Several methods are being evolved to construct high-density SNP maps of the human genome as opposed to the formerly used gel based resequencing technique. In this regard, the use of microarray technology could prove to be a nifty means of developing such maps.

Tang *et al* (1999) successfully detected such genetic variations using silicon based genomic DNA arrays, and chose the Human Platelet Alloantigens (HPA-1, HPA-2 and HPA-4) for genotyping. These antigens are implicated in a wide spectrum of autoimmune diseases such as neonatal alloimmunethrombocytopenia, posttransfusion purpura and refractoriness to platelet transfusions and could act as an excellent tool to investigate the utility of chip-based diagnostics. For their study, single-stranded DNA templates were covalently bonded to silicon chip wells using the N-Succinimidyl (4-iodoacetyl) aminobenzoate chemistry (SIAB). In addition to this, a mixture of selected dNTPs and ddNTPs, PROBE (Primer-Oligo base extension) primer, and thermosequenase were added. The chip was then placed in a chamber made from a microscope slide and seal film that was then kept in an *in situ* PCR machine and the PROBE reaction commenced. Thereafter, following the removal of enzyme and buffer salts by washing, a matrix solution of 3-hydroxypicolinic acid was dispensed into each well using a piezoelectric pipette. The chip was then transferred into a Mass spectrometer and analyzed using MALDI-MS (Matrix assisted laser desorption ionization mass spectrometry). Results indicated an unambiguous detection of heterozygotes and homozygotes in addition to which the use of such a

miniaturized reaction chamber greatly reduced the sample volume required for analysis and hence the cost analysis therein.

Several investigations analogous to the pioneering *Saccharomyces* experiment discussed earlier attempt at identifying disease genes and evaluating their expression using microarray technology. Whitney *et al* (1999) attempted to identify probable causative gene(s) involved in the onset of Multiple Sclerosis (MS). This is a disease of the CNS in which there is inflammation and damage to the myelin tissue resulting in errors in neural transmission and grave damage to the brain and spinal chord. In their study the group using microarrays of over 5,000 genes, compared the gene expression profile of normal and MS lesion white matter from the same patient. Sixty-two differentially expressed genes, identified with a wide variety of functions in transport, immune related functions, and signaling and myelin formation, included among others the Duffy chemokine receptor, interferon regulatory factor-2, and tumour necrosis factor and alpha receptor-2. This suggests a putative involvement of these genes in the onset of MS.

Aside from the sequencing of the human genome and construction of maps of all markers, the initiation of microarray technology has also enabled a clarified vision of the genome expression circuitry. Gene-specific transcription activators are among the main factors that specifically shape the transcriptome (transcription unit) profiles. Microarray experiments were devised that addressed the role of the transcription factors in the genome-wide expression profile. The first genome to be studied in this manner was the yeast (*S.cerevisiae*) genome and its conclusion enabled the characterization of connections between the different yeast regulatory networks (Devaux *et al*, 2001). The data collected on this investigation is also available through an on-line database, yTAFNET (<http://transcriptome.ens.fr/ytafnet/>). Besides studies on transcriptomes, efforts on investigating the expression and interaction between proteins in diverse biochemical activities are also underway. For instance, Zhu *et al* (2001) formed a high spatial density microarray of the yeast proteome by cloning 5800 open reading frames and overexpressing and purifying their corresponding proteins. The proteins were then printed onto slides and screened for their ability to interact with proteins and phospholipids. In this manner, they were able to identify many new calmodulin and phospholipid-interacting proteins and also a common potential

binding motif for many of the calmodulin-binding proteins.

Cancer Research

By far, the most exhaustive use of microarrays has been in cancer research and therapy. The aberrant proliferation of cells in cancer ostensibly occurs due to either a functional failure of tumor suppressor gene products in controlling cell division or due to an activation of oncogenes. Either way following the initiation of the initial tumorigenic mutation, there could be several genes whose products may play vital roles in propelling the cell towards malignancy. The application of microarray has not only enabled a more discerning perspective about the genetic origins of different cancers but also facilitated prognosis and aided in the designing of possible countermeasures to this dreaded disease.

One of the most common mechanisms of oncogene activation is gene amplification. Heiskanen *et al* (2000) developed cDNA microarrays to identify amplified genes and monitor their expression levels in cancer. In their study, they generated microarrays containing 14 different cDNAs representing genes known to be amplified in cancer cell lines such as *cdk4*, *mdm2*, *os4*, *os9*, *mycn*, *myc*, etc. In addition, microarrays of nearly 1400 cDNAs were also used to screen for unknown amplified genes. These arrays were then exposed to total genomic DNA that had been labelled with biotin by nick translation and the fluorescence detected using a confocal scanning microscope. Amplified genes were detected according to the distribution of signal intensities and a fold change in signal intensity was calculated by dividing the intensity of any spot by the average intensity of the non-amplified spots on the array. Results indicated a detection of 11 of the known gene amplifications, with the amplification level ranging from 5-fold (ERBB2) to 100-fold (MYCN). Additionally the intensity signal of a previously unidentified sequence tag representing ZNF133 was observed to have increased nearly 8-fold suggesting gene amplification. ZNF133 had been mapped earlier to 20p11.2 by *in situ* hybridization. In this manner microarrays had been employed successfully to perform both copy number and expression analysis.

Welsh *et al* (2001) used microarrays to identify possible genetic markers of epithelial ovarian cancer and constructed oligonucleotide arrays complementary to more than 6000 genes whose expression they planned to detect and correlate within aggregate normal and malignant tissues. For this, an "Array of

Arrays" was developed consisting of single glass wafer that contained 49 individual arrays separated by gaskets within a custom built chamber. RNA was isolated from 27 flash frozen serous papillary adenocarcinomas of the ovary and three normal samples of whole ovarian tissue that had been selectively macro-dissected into epithelial and stromal fractions. RNA from endothelial and activated B cells was also isolated to distinguish patterns of gene expression consistent with presence of blood vessels and/or infiltrating immune cells. Following RT PCR and labeling of the cDNA, hybridization was carried out and fluorescing hybrids detected using a confocal microscope. Softwares CLUSTER and TREEVIEW were used to select group and visualize genes whose expression varied across the samples. To identify potential tumor markers, the hybridization intensity of each gene, in normal as opposed to malignant tissue samples, was compared and the genes ranked accordingly. Hierarchical clustering of the expression data revealed distinct groups of samples.

Normal tissues were readily distinguishable from tumour tissues and tumours could also be divided into major groups analogous to the histological and clinical findings. Over 100 genes were found to be highly expressed in normal tissue and underexpressed in the tumors. These included the genes, c-fos, jun-B and EGR-1. Another cluster profile represented several smaller clusters of genes that were under expressed in normal tissues but overexpressed in tumor tissues. These included the genes such as HE4, the Preferentially Expressed Antigen of the Melanoma gene (PRAME) and c-erb-B2. A very high expression of a very large group of ribosomal genes was also noted suggesting a higher metabolic rate of the cancer cells. A proliferative cluster was also identified that showed increased expression of the genes connected with the cell cycle such as *cdc28*, protein kinases 1 and 2, *cdc25B* and *cdc20* indicating a possible correlation with aggressive tumour behavior. Finally after ranking the genes based on their expression based intensity, they proposed CD24, HE4 a secreted protease inhibitor, CD9, tumor associated antigen GA733-2, cytokeratins 7,8,18,19 and MUC-1 to be possible diagnostic markers for epithelial ovarian cancers, since they were most highly expressed in these cells.

Su *et al* (2000) applied cDNA microarrays to recognize probable Tumour Suppressor genes by comparing the expression profiles of the human melanoma cell lines, UACC903, UACC903 (+6) and

SRS3. They compared the gene expression profile of nearly 3317 genes between these three cell lines in pairs thereby identifying genes with tumour suppressor activities. Of the 3317 genes studied, about 321 showed expression changes between at least one pair of the three cell lines. Interestingly about 12 genes displayed higher levels of expression in UACC903 (+6) than the other two cell lines indicating presence of candidate tumour suppressor genes. These genes included that of Cx43 (connexin gene), monocyte chemotactic protein 1 and cysteine proteinase P32alpha. TP53 mutations are one of the commonest genetic alterations reported in human malignancies. Around 570 different TP53 mutations have been identified so far using the time bound traditional methods of SSCP/Heteroduplex/DGGE in combination with sequencing. Considering the laboriousness and the considerable time taken by these methods in the detection of mutations, it is fathomed that the entire mutation spectrum of the TP53 gene is still incomplete.

Wen *et al* (2000) compared the efficacy of using microarrays in the detection of mutations in the TP53 gene as opposed to the conventional method of SSCP followed by DNA sequencing. They analyzed genomic DNA from 108 ovarian epithelial carcinomas for mutations in the p53 gene using the SSCP method. They also detected 54 mutant shifts manual gel based sequencing confirmed mutations in 53 of the samples. The remaining tumour DNA samples that did not indicate a 'mobility shift' during SSCP were also subjected to sequencing, from which mutations were detected in ten samples. They then analyzed tumour DNA from all 108 specimens using a p53 GeneChip Assay. For this, PCR products were labelled and washed over the microarray and allowed to hybridize with the complementary oligonucleotides. Fluorescing hybrids arrays were then read using the Gene Array Scanner. The results obtained indicate that the Gene Chip had detected 71 mutations as opposed to the 63 detected by traditional methods! Thus, in terms of accuracy, sensitivity and specificity, the use of microarrays presented a more viable option in the detection of mutations aspect that would be of tremendous prognostic use.

Microarrays may also be used in the grading of tumours, thereby improving the accuracy of prognosis. This was demonstrated by Alizadeh *et al* (2000), who used cDNA microarrays to characterize gene expression in B cell malignancies and correlate the information with the differentiation state of the

tumour. For their investigation, they designed a special microarray unit termed the "Lymphochip" that represented an array of all genes preferentially expressed in lymphoid cells and also genes with known or suspected roles in immunology or cancer. From the results it may be surmised that a survey of global gene expression in tumours using microarrays would enable the correlation of molecular heterogeneity of tumours with the severity of outcome of the disease an aspect that would enable the conclusion of an appropriate prognosis.

Microarray analysis of gene expression could also be used in predicting the benefit proffered by chemotherapeutic treatment of different cancers. On the basis of the gene expression profile using cDNA microarrays, Martin *et al* (2000) were first to identify genes with a marked change in expression levels in tumour cells as compared to normal cells. Secondly they were also able to characterize breast tumours into three different groups depending on their Estrogen Receptor status, tumor stage and tumor size. Based on which information, responsiveness to anti hormone therapy as well as prognosis of clinical outcome of the tumour could be adjudged. In their experiment, Martin *et al* used cDNA arrays to screen for over 7000 genes for their differential expression in tumour tissue as compared to normal. From this they were able to identify a set of 170 genes that showed a disparate expression profile in tumour cells as opposed to normal cells. Nearly 70% of the genes that were downregulated in the tumor cells were categorized as filamentous, cell surface genes with roles in adhesion, communication and maintenance of cell shape. Over 75% of the genes that showed an up-regulated expression in tumour cells were those involved in metabolism, macromolecular synthesis and disruption of the extracellular matrix. They studied mRNA expression pattern of these genes and prepared arrays of over 124 different genes by spotting PCR products onto a nylon membrane. These gene tags included 89 normal cell specific genes and 18 tumor cell specific genes as well as previously reported cancer genes and housekeeping genes. cDNA probes were then generated from breast tumour tissue samples from 18 patients and seven breast cell lines. Following hybridization, cluster analysis was used to group those genes with similar expression patterns. Results indicated expression of two clusters of genes represented by p53 and maspin to be strongly correlated with estrogen receptor status. A third cluster that included the HSP-90 proved to be

instructive in relation to the stage of the cancer. A fourth gene cluster including the gene for Keratin proved to be an informative predictor of tumour size. Using these expression profiles, they were able to identify those patients who would possibly be unresponsive to anti hormonal Tamoxifen based therapy. They were also able to correlate gene expression with the severity of the tumour.

Scherf *et al* (2000) initiated the first investigation to generate databases that combined gene expression studies with molecular pharmacology. They used cDNA microarrays to assess the gene expression profiles in 60 human cancer cell lines. Clustering the cell lines depending on gene expression patterns in relation to drug response for the clinical agents, 5-fluorouracil and L-asparaginase, they were able to correlate variations in transcript levels with drug sensitivity and resistance. This is a very promising development for pharmacological research and holds tremendous potential for exploiting the use of novel drugs in combating cancer.

Pharmacological Screening and Toxicology Studies

Extract of *Ginkgo biloba* leaves is a commonly prescribed herbal medicine for chronic, age-related neurological disorders such as Alzheimer's disease or various common geriatric complaints including vertigo, depression, short-term memory loss, hearing loss, lack of attention or vigilance. Watanabe *et al* (2001) used cDNA microarrays to identify the *in vivo* neuromodulatory effects of the *Ginkgo biloba*. In their study, among 20 female adult C57BL6 mice, 10 were fed on a diet supplemented with ginkgo leaf extracts and the remaining 10 mice were taken as controls and kept on a low-flavonoid maintenance diet. Microarrays representing all functionally characterized sequences (6,000) as well as 6,000 expressed sequence tag (EST) clusters were constructed. After four weeks on the prescribed diet hippocampi and cortices were dissected from these animals, from which total RNA was extracted and processed to obtain biotinylated cRNA. Following hybridization, expression levels of the 12,000 genes were assessed. About 10 genes were up regulated in expression level by 3-fold or more as confirmed by RT-PCR. These genes included transthyretin from the hippocampus and nine genes in the cortex including the AMPA-2 channel, neuronal tyrosine/threonine phosphatase 1, and microtubule-associated τ , all of which were surmised to have neuroprotective roles. Ion channels, growth hormones, and transcription factors were also

among those genes whose expression was enhanced by diet supplementation with Ginkgo. In this manner the employment of microarrays clarified the role of the bioactive component of the medication in its contribution to the palliative effect overall. This has unmistakable implications in drug research.

High-density microarrays can also be useful tools in identifying mechanisms of toxicity in response to the action of drugs and chemicals. Bulera *et al* (2001) administered 6 hepatotoxicants known to induce adverse liver effects through different mechanisms namely, microcystin-LR (MLR), phenobarbital (PB), lipopolysaccharide (LPS), carbon tetrachloride (CT), thioacetamide (THA), and cyproterone acetate (CPA) to groups of male Wistar rats. Following which liver mRNA was isolated and used to generate biotinylated cRNA for hybridization to a custom 1,600-rat gene DNA microarray. Hybridization data was analyzed using a cluster-based analysis and the experimental liver toxicity from the different treatments confirmed by means of concurrent histopathology, liver enzymes and bilirubin assays. This toxico-genomic analysis identified several clusters of genes to be affected by the hepatotoxicants under study. This proved that high-density microarray expression data was not only useful in identifying the groups of genes involved in toxicity but also in determining or predicting toxic liver effects.

Study of Multistage Developmental Processes

Microarrays have also been successfully used in studying gene expression patterns during several multistage developmental processes. Liu *et al* (2001) employed cDNA microarrays to investigate the gene expression and regulation involved in folliculogenesis and was able to identify genes with stage-specific roles in follicular development.

Microarrays—the Negative Side

The technology being still in its early stages of development, problems are often encountered during its deployment. However, the advantages outweigh the difficulties and particularly when seen in the context of overwhelming information to be gained these problems should be considered as challenges rather than obstacles (www.ipam.ucla.edu/programs/fg2000/pa.html).

The primary stumbling block is the cost factor. A routine chip can cost anything from US \$ 1100 to 200,000. The cost of a microarray unit is dependent on the gene density on the chip; the nature of substrate used (silica/silicon/nylon/glass) and the type

of molecule placed on the unit (cDNA/mRNA/genomic DNA). What would be a worthwhile investment to any major pharmaceutical or biotechnological company might stretch the resources of an ordinary laboratory especially in the Indian scenario. However prices are eventually expected to fall as usage and number of users increase. Presently, there are several websites and links to sites offering free and complete information on microarray experiment and analysis such as, <http://cmgm.stanford.edu/pbrown>, <http://www.bsi.vt.edu/ralscher/gridit> and <http://ihome.cuhk.edu.hk>, etc.

Essentially in a hybridization technology, some of the problems encountered under experimental conditions are cross-hybridization and strong background signals that may result in the wrongful detection of false positives. Additionally problems may also arise due to the two commonly used fluorescent dyes Cy3 and Cy5. There has been a bias observed in the incorporation of these dyes due to their chemical structure that consequently could result in false positive signals. However, these problems may be tackled by labeling controls and samples with either dye in duplicate experiments. A novel use of an amino allele compound has been found to cause the dyes to be incorporated at a greater and more equal amount, thereby reducing bias and increasing sensitivity. However, a more nagging problem that needs to be tackled is differential expression of genes, such as those of some hormones, that function at extremely low levels and specific in their expression, cannot be detected.

Possibly the greatest challenges to array technology lie in the analysis of the gene expression data. Especially due to the multifarious dimensionality of the data obtained, there are many challenging problems of multiplicity and multivariate analysis that needs to be addressed. The issues need to be tackled include the development of better image analysis of the fluorescent signals, improvement of design of oligonucleotides for purposes of genotyping, analysis of expression data including classification of cells based on expression pattern and diagnostic disease classification systems. For the present however, free software available online provides basic microarray data storage, retrieval and analysis capabilities to an average laboratory. Some of these websites giving information include <http://www.microarrays.org/software.htm> that gives the AMAD software package. Other packages including AMADA, arraySCOUT, CLEAVER, clusFAVOR, CLUSTER, EXPRESSION

PROFILER by the European Bioinformatics Institute (EBI), etc., are also available as free downloads for analysis (<http://linkage.rockefeller.edu/wli/microarray/soft.html>).

Conclusion

The advent of microarray technology has provided the molecular biologist with a spectacular opportunity to visualize gene expression. It is now possible to study the interaction of several genes in disease and normal and also obtain a better standpoint about the underlying molecular proceedings. Aside from the obvious applications in terms of annotating different genomes, identification of culpable genes, pharmacogenetics and drug testing, microarrays can be exploited to piece together virtually any complex biological jigsaw puzzle. Ongoing studies attempt to investigate diverse subjects such as the genetic basis of aging and the management of gerontological medicine (Helmberg, 2001; Park *et al.*, 2001) to the complex interaction between a microbial pathogen and the clinical management of infectious disease using microarrays (Cummings & Relman, 2000).

Despite problems of properly evaluating and managing the data overload generated by the use of microarrays, biotechnologists have realized the awesome promise of this brilliant innovation and are working around these 'teething troubles' to ensure a more coherent approach at extracting information in the quickest possible time. In the words of Stephen Fodor "The potential for this technology is enormous and only limited by imagination". Indeed the future has arrived at our doorstep and it is now our initiative to throw our gates wide open and give it the welcome it deserves.

References

- Alizadeh A A *et al.*, 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature (Lond)*, **405**, 503-511.
- Beaucage S L, 2001. Strategies in the preparation of DNA oligonucleotide arrays for diagnostic applications. *Curr Med Chem*, **8**, 1213.
- Brown P O & Botstein D, 1999. Exploring the New World of the genome with DNA microarrays. *Nat Genet*, **21**(suppl 1), 33-37.
- Bulera S J *et al.*, 2001. RNA expression in the early characterization of hepatotoxins in Wistar rats by high-density DNA microarrays. *Hepatology*, **33**, 1239-1258.
- Cheung V G *et al.*, 1999. Making and reading microarrays. *Nat Genet*, **21**(Suppl 1), 15-19.
- Cummings C A & Relman D A, 2000. Using DNA microarrays to study host-microbe interactions. *Emerg Infect Dis*, **6**, 513.
- DeRisi J L *et al.*, 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, **278**, 680-686.

- Devaux F *et al*, 2001. Transcriptomes, transcription activators and microarrays. *FEBS Lett*, **498**, 140-144.
- Duggan D J *et al*, 1999. Expression profiling using cDNA microarrays. *Nat Genet*, **21**, 10-14.
- Dunham I *et al*, 1999. The DNA sequence of human chromosome 22. *Nature (Lond)*, **402**, 489-495.
- Eisen M B *et al*, 1998. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA*, **95**, 14863-14868.
- Ekins R & Chu F W, 1999. Microarrays: their origins and applications. *Trends Biotechnol*, **17**, 217-218.
- Fodor S P *et al*, 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science*, **251**, 767-773.
- Functional genomics, www.ipam.ucla.edu/programs/fg2000/pa.html.
- Guigo R *et al*, 2000. An assessment of gene prediction accuracy in large DNA sequences. *Genome Res*, **10**, 1631-1642.
- Heiskanen M A, 2000. Detection of gene amplification by genomic hybridization to cDNA microarrays. *Cancer Res*, **60**, 799-802.
- Helmberg A, 2001. DNA-microarrays: Novel techniques to study aging and guide gerontologic medicine. *Exp Gerontol*, **36**, 1189-1198.
- <http://cmgm.stanford.edu/pbrown/mguide/index.html>.
- Leming Shi. DNA Microarray (Genome Chip) - Monitoring the Genome on a Chip. <http://www.gene-chips.com>.
- Lipshutz R J, 1999. High density synthetic oligonucleotide arrays. *Nat Genet*, **21**(Suppl 1), 20-24.
- Liu H C *et al*, 2001. Application of complementary DNA microarray (DNA chip) technology in the study of gene expression profiles during folliculogenesis. *Fertil Steril*, **75**, 947-955.
- Martin, K J *et al* 2000. Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res*, **60**, 2232-2238.
- Park W Y, 2001. Gene profile of replicative senescence is different from progeria or elderly donor. *Biochem Biophys Res Commun*, **282**, 934-939.
- Scherf U *et al*, 2000. A gene expression database for the molecular pharmacology of cancer. *Nat Genet*, **24**, 236-244.
- Shoemaker D D *et al*, 2001. Experimental annotation of the human genome using microarray technology. *Nature (Lond)*, **409**, 922-927.
- Su Y A *et al*, 2000. Identification of tumor-suppressor genes using human melanoma cell lines UACC903, UACC903 (+6), and SRS3 by comparison of expression profiles. *Mol Carcinog*, **28**, 119-127.
- Tang K *et al*, 1999. Chip-based genotyping by mass spectrometry. *Proc Natl Acad Sci USA*, **96**, 10016-10020.
- Watanabe Coran M. H *et al*, 2001. The *in vivo* neuromodulatory effects of the herbal medicine *Ginkgo biloba*. *Proc Natl Acad Sci USA*, **98**, 6577-6580 [epub ahead of print].
- Welsh, J B *et al*, 2001. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc Natl Acad Sci USA*, **98**, 1176-1181.
- Wen W H *et al* 2000. Comparison of TP53 mutations identified by oligonucleotide microarray and conventional DNA sequence analysis. *Cancer Res*, **60**, 2716-2722.
- Whitney LW *et al*, 1999. Analysis of gene expression in multiple sclerosis lesions using cDNA microarrays. *Ann Neurol*, **46**, 425-428.
- Zhu H *et al*, 2001. Global Analysis of Protein Activities Using Proteome Chips. *Science*, July 26 [epub ahead of print].