Microarray

Chapter · October 2010

DOI: 10.1007/978-90-481-3261-4_13

3 authors:

Mousumi Debnath
Manipal University Jaipur
80 PUBLICATIONS **1,113** CITATIONS
SEE PROFILE

GBKS Prasad
Jiwaji University
277 PUBLICATIONS **2,493** CITATIONS
SEE PROFILE

Prakash S Bisen
Jiwaji University
561 PUBLICATIONS **5,868** CITATIONS
SEE PROFILE

Debnath Mousumi, Prasad GBKS, Bisen PS. Molecular Diagnostics: Promises and Possibilities Dordrech Heidelberg London, Springer, 2010 pp 193-208.

CHAPTER 13

MICROARRAY

Abstract

DNA microarrays are the modern, parallel version of classic molecular biology hybridization techniques allowing the monitoring the expression of thousands of genes simultaneously in a single run. Microarray allows rapid measurement and visualisation of differential expression of the whole genome scale. Using microarrays , RNA sample from any cell or tissue type can be analysed for changes in transcript levels. Such changes are characteristics of developmental processes , cancer , infection , responses to drugs , sequencing and genotyping . The technique permits the complete analysis of genetic material and the monitoring of expression changes occurring in a biological sample under various conditions. The application of microarrays will have a significant role in clinical and basic science. Microarrays have been used successfully in various research areas including sequencing, single nucleotide polymorphism detection, characterization of protein-DNA interactions, DNA computing, mRNA profiling, and many more. Applications of microarrays also include gene expression studies, disease diagnosis, pharmacogenomics, drug screening, pathogen detection, and genotyping. Improved understanding of the gene changes occurring in human diseases will lead to the development of new diagnostic tools as well as more effective drugs.

Key words

Gene expression, profiling, microarray, molecular medicine, drug screening, hybridisation , single nucleotide polymorphism , transcriptional factors , expression of genes ,SNP, genotyping ,

13. 1 PROLOGUE

It is widely believed that thousands of genes and their products in a given living organism function in a complicated and orchestrated way that creates the mystery of life. However, traditional methods in molecular biology generally work on a “one gene one experiment” basis which means that the throughput is very limited and the whole picture of gene function is hard to obtain. Another important and interesting question in biology is how gene expression is switched on and off and how genes are regulated. Since almost all cells in a particular organism have an identical genome, differences in gene expression and not the genome content are responsible for cell differentiation during the life of the organism.

Gene regulation in eukaryotes is not well understood, but there is evidence that an important role is played by a type of proteins called *transcription factors*. The transcription factors can attach (bind) to specific parts of the DNA, called transcription factor *binding sites* (i.e., specific, relatively short combinations of A, T, C or G),

which are located in so-called *promoter* regions. Specific promoters are associated with particular genes and are generally not too far from the respective genes, though some regulatory effects can be located as far as 30,000 bases away, which make the definition of the promoter difficult. Transcription factors control gene expression by binding the gene's promoter and either activating (switching on) the gene's transcription, or repressing it (switching it off). Transcription factors are gene products themselves, and therefore in turn can be controlled by other transcription factors. Transcription factors can control many genes, and some genes are controlled by combinations of transcription factors. Understanding, describing and modelling such gene regulation networks are one of the most challenging problems in functional genomics.

There are over 200 known transcription factors in yeast, over 600 in worm and fly, and over 1500 in weed, but the real numbers are probably higher, since more than half of the predicted genes in these organism have unknown function and there are transcription factors likely to be found among the unknown genes. In addition, transcription factors are not the only proteins participating in gene regulation. Proteins affecting chromatin structure are another example, and it is known that at least some regulation is happening in the translation stage.

The complete sequencing of several genomes, including that of the human, has signaled the beginning of a new era in which scientists are becoming increasingly interested in functional genomics; that is, uncovering both the functional roles of different genes, and how these genes interact with, and/or influence, each other. Increasingly, this question is being addressed through the simultaneous analysis of hundreds to thousands of unique genetic elements with microarrays. Microarrays and computational methods are playing a major role in attempts to reverse engineer gene networks from various observations (Kothapalli *et al.,* 2002). Note that in reality the gene regulation is likely to be a stochastic and not a deterministic process. Already, analytical strategies have subdivided into distinct 'omic' domains, such as genomics, proteomics, and metabolomics. This enables researchers to examine not only genetic elements, but also the corresponding proteins and metabolites derived from these genes. All 'omic' technologies share the need for fresh, innovative looks at data analysis. To date, transcriptomics is the most widely studied molecular approach, enabling researchers to examine subtle differences in thousands of mRNA levels between experimental samples, medical biopsies, etc. Although mRNA is not the end product of a gene, the transcription of a gene is both critical and highly regulated, thereby providing an ideal point of investigation (Brazma and Vilo , 2000). Development of microarrays has permitted global measurement of gene expression at the transcript level and provided a glimpse into the coordinated control and interactions between genes .

13.2 THE CONCEPT OF MICROARRAY

The concept of using microarrays can be traced back 25 years to the introduction of the Southern blot. Modern microarrays analysis was introduced in 1995 by a Stanford University research team led Pat Brown and Ron davis . A microarray is an analytical device that comprises an array of molecules (oligonucleotides, cDNAs, clones, PCR products, polypeptides, antibodies, and others) or tissue sections immobilized at discrete ordered or nonordered micrometer-to-millimeter-sized locations on the surface of a porous or nonporous insoluble solid

support. Microarrays exploit the preferential binding of complementary single-stranded nucleic acid sequences(Barrett *et al.,* 2003). They are miniaturized biological devices consisting in molecules , for examples DNA or protein, named probes, that are orderly arranged at a microscopic scale (Fig 13.1) onto a solid support such as a membrane or a glass microscopic slide .A microarray is typically a glass (or some other material) slide, on to which DNA molecules are attached at fixed locations (spots). There may be tens of thousands of spots on an array, each containing a huge number of identical DNA molecules (or fragments of identical molecules), of lengths from twenty to hundreds of nucleotides. For gene expression studies, each of these molecules ideally should identify one gene or one exon in the genome, however, in practice this is not always so simple and may not even be generally possible due to families of similar genes in a genome. Microarrays that contain all of the about 6000 genes of the yeast genome have been available since 1997. The spots are either printed on the microarrays by a robot, or synthesized by photo-lithography (similarly as in computer chip productions) or by ink-jet printing(Hughes *et al.,* 2001). The light-generated oligonucleotide array, developed by Affymetrix, Inc. (Santa Clara, CA), involves synthesizing short 25-mer oligonucleotide probes directly onto a glass slide using photolithographic masks. Sample processing includes the production of labeled cRNA, hybridization to a microarray, and quantification of the obtained signal after laser scanning. Regardless of the array used, the output can be readily transferred to commercially available data analysis programs for the selection and clustering of significantly modified genes(Culhane *et al.,* 2003).

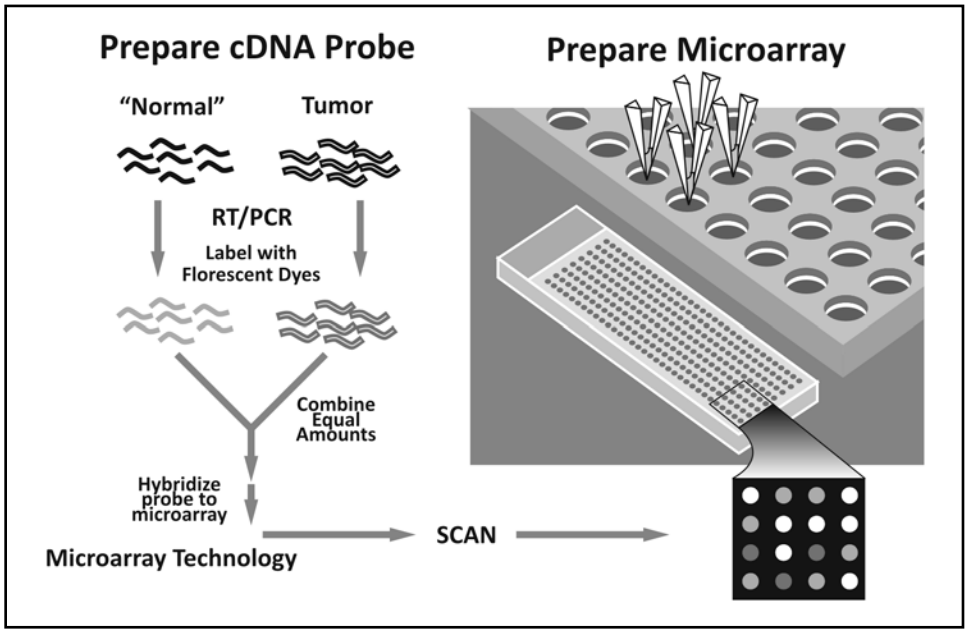

Fig 13.1 Concept of microarray

DNA microarrays are widely used to analyze genome-wide gene expression patterns and to study genotypic variations. Before a nucleic acid (target) can be used for hybridization to the probes of a microarray, it needs to be extracted from the tissue and labeled. The total mRNA from the cells in two different conditions (Kuo *et al.,* 2002) is extracted and labeled with two different fluorescent labels: for example a green dye for cells at condition 1 and a red dye for cells at condition 2 (to be more accurate, the labeling is typically done by synthesizing single stranded

DNA's that are complementary to the extracted mRNA by a enzyme called reverse transcriptase). Frequently, it also needs to be amplified to increase detection sensitivity. Both extracts are washed over the microarray. Labeled gene products from the extracts hybridize to their complementary sequences in the spots due to the preferential binding - complementary single stranded nucleic acid sequences tend to attract to each other and the longer the complimentary parts, the stronger the attraction .To increase throughput and minimize costs without reducing gene expression different fluorescent dyes namely Alexa 488, Alexa 594, Cyanine 3 and Cyanine 5 can also be used for hybridization.

During a hybridization process, labeled target molecules with sequences complementary to the probes are captured quantitatively. Subsequently, a reader measures the amount of label on each probe. To generate accurate and informative data, one of the most critical aspects of these experiments is the quality of both the isolated and the labeled nucleic acid samples (Lockhart *et al.*,1996).

The dyes enable the amount of sample bound to a spot to be measured by the level of fluorescence emitted when it is excited by a laser. If the RNA from the sample in condition 1 is in abundance, the spot will be green, if the RNA from the sample in condition 2 is in abundance, it will be red. If both are equal, the spot will be yellow, while if neither is present it will not fluoresce and appear black. Thus, from the fluorescence intensities and colours for each spot, the relative expression levels of the genes in both samples can be estimated.

The raw data that are produced from microarray experiments are the hybridised microarray images. To obtain information about gene expression levels, these images should be analysed, each spot on the array identified, its intensity measured and compared to the background. This is called image quantitation.

Microarrays are miniature arrays of gene fragments attached to glass slides. Monitoring the expression of thousands of genes simultaneously is possible with DNA microarray analysis. These devices have been highly effective for the simultaneous detection of large numbers of analytes in a sample, and microarrays have quickly emerged as important analytical tools in many branches of the biological sciences. A microarray-based analytical strategy is quicker and more convenient than serial testing for each analyte, and it has been successfully applied to both immunoassays and DNA-based assays. The current scope of microarray applications includes sequencing by hybridization, resequencing, mutation detection, assessment of gene copy number, comparative genome hybridization, drug discovery, expression analysis, and immunoassay (protein microarrays). In addition, oligonucleotide microarrays have been used for a nonbiological application computing (Hacia, 1999).

The presentation of thousands of gene fragments in a single array allows investigators to define changes in gene expression within a significant fraction of the total genome. For example, cancers will no longer be classified on the basis of a few proteins, but will be characterized by the change in hundreds of specific genes. Animal models will be related to human diseases not by appearance alone but by the changes that occur in thousands of genes. By coupling microarray analysis with the results from genome sequencing projects, researchers can analyze the global patterns of gene expression of an organism during specific physiological responses or developmental processes.

Presently, two technologies dominate the field of high-density microarrays: cDNA arrays and oligonucleotide arrays(Kane *et al.,* 2000). The cDNA array has a long history of development stemming from immunodiagnostic work in the 1980s; however, it has been most widely developed in recent years by Stanford University (California) researchers depositing cDNA tags onto glass slides, or chips, with precise robotic printers(Lipshutz *et al.,* 1999). Labeled cDNA fragments are then hybridized to the tags on the chip, scanned, and differences in mRNA between samples identified and visualized using a variation of the red/green matrix. A tissue microarray using paraffin embedded formalin fixed tissues can be used to evaluate the immunoprofiles of a large set of uterine adenocarcinomas with an extended panel of antibodies ,comparing the profile of primary cervical and andometrial adenocarcinomas .

Differentially expressed genes will be defined herein as gene data determined to be statistical outliers from some standard state, and which can not be ascribed to chance or natural variabilty (Li and Wong , 2001). Various creative techniques have been proposed and implemented for the selection of differentially expressed genes; however, none have yet gained widespread acceptance for microarray analysis (Kane *et al.,* 2000). Despite this, there remains a great impulse to develop new data analysis techniques, partly driven by the obvious need to move beyond setting simple fold change cut-offs which are out of context with the rest of the experimental and biological data at hand .This has been the case for many studies, where the selection of differential gene expression is performed through a simple fold change cut-off, typically between 1.8 and 3.0. There is an inherent problem with this selection criterion, as genes of low absolute expression have a greater inherent error in their measured levels. These genes will then tend to numerically meet any given fold change cut-off even if the gene is not truly differentially expressed. The inverse also holds true, where highly expressed genes, having less error in their measured levels, may not meet an arbitrary fold-change cut-off of 2.0 even when they are truly differentially expressed. Therefore, selecting differentially regulated genes based only on a single fold change across the entire range of experimental data preferentially selects lowly expressed genes. This commonly used approach does not accommodate for background noise, variability, non-specific binding, or low copy numbers, characteristics typical of microarray data which may not be homogeneously distributed. Other approaches entail the use of standard statistical measures such as a student's *t*-test or ANOVA for every individual gene(Johnson and Wichern,2002). However, due to the cost of repeating microarray experiments, the number of replicates usually remains low, leading to inaccurate estimates of variance (Baldi *et al.,* 2001). Furthermore, due to the low number of replicates, the power of these "gene-by-gene" statistical tests to differentiate between regulated and non-regulated genes also remains very low.

Protein microarrays are versatile tools for parallel, miniaturized screening of binding events involving large numbers of immobilized proteins in a time- and cost-effective manner. They are increasingly applied for high-throughput protein analyses in many research areas, such as protein interactions, expression profiling and target discovery. While conventionally made by the spotting of purified proteins, recent advances in technology have made it possible to produce protein microarrays through in situ cell-free synthesis directly from corresponding DNA arrays (Stoevesandt *et al.,*2009).

13.3 CURRENT CHALLENGES OF MICROARRAYS

DNA microarray technology is currently an area of great interest (Fig 13.2). Also called "genechip" technology, it incorporates molecular genetics and computer science on a massive scale. This technology can rapidly provide a detailed view of the simultaneous expression of entire genomes and provide new insights into gene function, disease pathophysiology, disease classification, and drug development (Steven and Greenberg, 2001).The application of microarrays will have a significant role in clinical and basic science. Improved understanding of the gene changes occurring in human diseases will lead to the development of new diagnostic tools as well as more effective drugs (Hacia ,1999). In the basic sciences, microarray technology will permit researchers to elucidate the genetic pathways that are followed in the development, aging, and pathology of specific cells or tissues of the body.

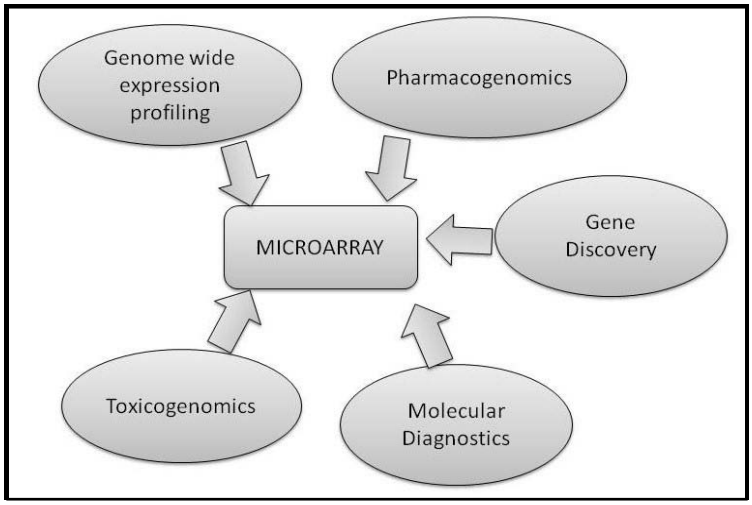

Fig 13.2 The wide applications of microarrays

Microarray technology will help researchers to learn more about many different diseases, including heart disease, mental illness and infectious diseases, to name only a few. One intense area of microarray research at the National Institutes of Health (NIH) is the study of cancer. In the past, scientists have classified different types of cancers based on the organs in which the tumors develop. With the help of microarray technology, however, they will be able to further classify these types of cancers based on the patterns of gene activity in the tumor cells. Researchers will then be able to design treatment strategies targeted directly to each specific type of cancer. Additionally, by examining the differences in gene activity between untreated and treated tumor cells - for example those that are radiated or oxygen-starved - scientists will understand exactly how different therapies affect tumors and be able to develop more effective treatments.

Within the last decade protein microarray technology has been successfully applied for the simultaneous identification, quantification and functional analysis of proteins in basic and applied proteome research. These miniaturized and parallelized assay systems have the potential to replace state-of-the-art singleplex analysis systems. However, prior to their general application in robust, reliable, routine and high-throughput applications it is mandatory that they demonstrate robustness, sensitivity, automation and appropriate pricing .The microarray analysis provides interesting leads in almost all fields of biology , offers a genome wide glimpse into genetic "terra incogniat," and challenges scientists to explore this unknown world .The biomedical community is developing new methods of data analysis to more efficiently process the massive data sets produced by microarray experiments. Systematic and global mathematical approaches that can be readily applied to a large number of experimental designs become fundamental to correctly handle the otherwise overwhelming data sets. The various aplications of DNA microarrays are discussed below in details .

13.3.1 GENE DISCOVERY

The use of high density DNA microarrays to identify sets of genes with similar expression pattern is rapidly becoming a wide spread approach patterns is rapidly becoming becoming a widespread approaches for understanding biological processes . Typically, microarray data is analyzed by hierarchical clustering, self-organizing maps, *K*-means clustering or principle component analysis. Most of these approaches readily identify clusters of tens to hundreds of genes that demonstrate similar expression patterns. One logical systematic approach to study a cluster of genes with similar expression profiles is to analyze the promoter sequences for each member of the gene clusters and attempt to identify transcription factors that might be crucial for regulating their expression .

Detecting and determining the relative abundance of diverse individual sequences in complex DNA samples is a recurring experimental challenge in genomes. DNA microarrays are utilized in the analysis of complex DNA samples. With the objective of discovering novel putative interventions sites for anticancer therapy, transcriptional profiles of breast cancer, lung squamous cell cancer (LSCC), lung adenocarcinoma (LAC), and renal cell cancer (RCC) were compared. They wanted to search for genes not only highly expressed in the majority of patient samples but which also exhibit very low or even absence of expression in a comprehensive panel of 16 critical (vital) normal tissues. In addition to already known tumor-associated genes, the majority of identified genes have not yet been brought into context with tumorigenesis such as genes involved in bone matrix mineralization (OSN, OPN, and OSF-2) in lung, breast, and kidney cancer or genes controlling Ca(2+) homeostasis (RCN1,CALCA, S100 protein family). EGLN3, which recently has been shown to be involved in regulation of hypoxia-inducible factor, was found to be highly up-regulated in all RCCs and in half of the LSCCs analyzed. Furthermore, 42 genes, the expression level of which correlated with the overall survival of breast cancer patients, were identified (Favis and Barany, 2000; Favia *et al.*, 2000; Brenton *et al.*, 2001).

DNA microarray analysis helped to identify several of those genes involved in several of these genes are involved in pathological processes relating to arthritis, including apoptosis, inflammation, and cellular proliferation.

One interesting gene, follistatin-like gene, is highly expressed along the margin of contact between inflammatory synovial pannus and eroding bone, suggesting a role in joint destruction. The identification of genes predisposing to human diseases is of paramount importance .Numerous genes may play a role in drug response and toxicity ,introducing a daunting level of complexicity into the search for candidate genes .Gene discovery in bladder cancer progression , gastrointestinal tract origin cancer cells have also been identified (Golub *et al.,* 1999).This technology has provided insights into gene function ,disease pathophysiology ,disease classification and and drug development. This technology is also called "genechip" technology as it incorporates molecular genetics and computer science on a massive scale (Steven and Green berg, 2001).

13.3.2 **GENE EXPRESSION PROFILING**

DNA microarray is widely used in gene expression analysis. Expression profiling by this technology permits the identification of genes underlying clinical heterogeneity of many diseases which might contribute to disease progression, thereby improving assessment of treatment and prediction of patient outcome. The gene expression profiles provide a molecular fingerprint of the transcriptome.

Microarray analysis has become a widely used tool for the generation of gene expression data on a genomic scale. The last several years have witnessed remarkable development of new microarray-based automated techniques allowing parallel analysis of multiple DNA samples. Two major current applications of DNA microarray**s** are gene expression profiling (Lockhart *et al.,* 1996 ; Mecham *et al.,* 2004) and gene mutation analysis (Hacia, *et al.,* 1996; Hakka, 1999). Gene expression profiling using DNA microarray**s** has already provided results that were not achievable a few years ago. For example, it allowed the molecular classification of leukemias (Golub, *et al.,* 1999) and other types of cancers (Brenton, *et al.,* 2001) and revelations about genetic network architecture (Tavazoie *et al.,* 1999).

Gene expression monitoring is not the only microarray application, another one is SNP detection. Microarray's are already producing massive amounts of data. These data, like genome sequence data, can help us to gain insights into underlying biological processes only if they are carefully recorded and stored in databases, where they can be queried, compared and analysed by different computer software programs.

Microarray expression analysis has become one of the most widely used functional genomics tools.Efficient application of this technique requires the development of robust and reproducible protocols.Lately optimisation of all aspects of the process, including PCR amplification of target cDNA clones, microarray printing, probe labeling, hybridization, including strategies for data normalization and analysis have been acheived. The 0.1% difference is exploited in DNA fingerprinting used in criminology. Variations in non-coding parts of the genome are analysed to produce patterns that can reliably distinguish individuals, an exception here is identical twins which are much harder to distinguish with DNA fingerprinting. Particularly important variations in individual genomes are the *single nucleotide polymorphisms* or SNPs, which can occur both in coding and non-coding parts of the genome. SNPs are DNA sequence variations which occur when a single base (A,C,G, or T) is altered so that

different individuals may have different letters in these positions. Particular nucleotides in SNP positions within or close to genes can influence the gene's protein product. Some 'abnormal' protein variants (mutant variants) are the cause of genetic diseases (Cargill *et al.,* 1999). SNPs may be responsible for many inheritable differences between individuals, and SNP variation may indicate the predisposition to a genetic disease. For example there is evidence that certain combinations of SNPs occur in individuals with Alzheimers disease. SNP analysis is therefore important for diagnostics and a SNP database is being built with the participation of the EBI. SNP analysis can also be used in population genetics, as some SNPs vary in frequency between populations. There are about 3 million SNPs collected in public SNP databases.

The sequencing projects have revealed that the genomes of what look like very different organisms may be quite similar. An example of this is similarity between yeast and worm described in the next section. It is estimated that the difference between human and chimpanzee genomes is only 1-3%, while between human and mouse only about 5% - 15%, depending on how the similarity is defined and measured (before the sequencing of chimp and mouse are well under way, these are only estimates). These similarities indicate close evolutionary relationships between these mammalian organisms. It is possible to build a *phylogenetic tree* of the evolution of proteins, genes and organisms based on sequence comparisons.

Gene annotation can be taken care to some extent of by links to sequence databases, unfortunately complicated many-to-many relationships between genes in the gene expression matrix and the features (spots) on the array makes it necessary to provide a full and detailed description of each feature on the array, as one gene can relate to several features on the array. The lack of standards in gene naming is another difficulty - a table relating each array feature present in the database to the list of all synonymous names of the respective gene is an essential part of a gene expression database. The ultimate goal of this work is to establish a standard for recording and reporting microarray-based gene expression data, which will in turn facilitate the establishment of databases and public repositories and enable the development of data analysis tools. Detecting and determining the relative abundance of diverse individual sequences in complex DNA samples is a recurring experimental challenge in analyzing genomes. This system should find numerous applications in genome-wide genetic mapping, physical mapping, and gene expression studies.

13.3.3 PHARMACOGENOMICS AND MICROARRAY

Pharmacogenomics deals with the genetic basis underlying variable drug response in individual patients. It encompasses sum of all genes. Numerous genes may play a role in drug response and toxicity. The newly emerging genomic technologies enable the search for relevant genes and their variants to include the entire genome and help in the search for candidate genes. Moreover pharmacogenomic analysis can identify the disease susceptibility genes representing potential new drug targets. DNA microarrays has alread revolutionised significant part of drug discovery and developmental process. It is rapidly becoming a fundamental tool in discovery based genomic and biomedical research (Wang *et al.,* 2003). DNA microarrays can be used to measure the expression patterns of

thousands of genes in parallel, generating clues to gene function that can help to identify appropriate targets for therapeutic intervention. They can also be used to monitor changes in gene expression in response to drug treatments . The DNA microarrays have proven to be a state of the art technique for high throughput comprehensive analysis of thousand of genes in parallel. The application of a DNA microarray to compare normal and pathological cells, tissues or organs may allow, along with classical positional cloning techniques, to speed up the discovery of genes and gene pathways implicated in several diseases. This in turn will result in further application of the DNA microarray technique in the field of pharmacogenomics in order to characterize and validate new therapeutic targets, their mechanism of action, metabolic pathways and unwanted secondary effects. However, the lack of standardized criteria for the analysis and interpretation of the huge amount of data generated by DNA microarrays may hamper its utilization on a routine basis in the human health domain.

In a recent study by Schmoock and his collegues (2009) a rapid oligonucleotide microarray assay based on genetic markers was succesfully performed for the accurate identification and differentiation of the agents of glanders and melioidosis *Burkholderia (B.) mallei* and *B. Pseudomallei*.Both the strains are potential agents of bioterrorism and cause infectious diseases of man and animals. They existing bacteriological and immunological methods of identification of *B. mallei* and *B. pseudomallei* are not efficient enough for the rapid diagnosis and typing of strains.

The current concept of drug therapy is to apply treatment for a large population group whereas application of pharmacogenomic will help individual application of drug therapy or on smaller patient subpopulation but whether this individualized medicine will lead to improved and economically feasible therapy is yet to be seen. This is again complicated by various ethical and legal issues .

With the advent of DNA microarray analysis, it is now possible to examine the response of virtually the entire human genome to cellular drug exposure and to uncover a wide variety of genes correlating with the establishment of drug resistance. Genechip arrays are now being used by pharmaceutical companies to improve nearly every aspect of the traditional drug discovery and developmental processs, including target identification, target validation, compound screening, lead optimization and clinical trials .The field of "pharmacogenomics" is likely to vastly increase our understanding of the mechanisms of drug action and how cells respond and adapt to drug exposure. The scientist hope to identify genes influencing drug response and ultimately tailor antihypersensitive therapy for individual patients .Microarrays and DNA analysis enable scientists to explore the whole genome and identify predictive markers of disease to drug response that may ultimately provide more tailored, effective, and safer courses of treatment to help avoid some of over 100,000 annual fatalities from adverse reactions in US alone. Many researchers agree that in DNA microarrays will be used as the most important diagnostic tools in near future.

However, DNA microarray studies typically result in the identification of hundreds of genes that may or may not be of relevance in vivo-particularly when large, genetically diverse study populations are used. The challenge to the researcher is to design experimental systems and approaches which minimize variability in the data, increase the reproducibility amongst experiments, allow array data from multiple experiments to be assessed by a variety of

statistical, supervised learning, and data clustering approaches, and provide a clear link between drug response and the expression of specific genes (Yuen *et al.,* 2002). Now the possibility to simultaneously analyze the expression of thousands of genes using DNA microarrays has allowed exploring the relationships between gene expression and sensitivity to several anticancer drugs. A number of studies using microarrays for identifying genes governing tumor chemosensitivity focused on tumor cell lines. Some clinical studies have also been carried out to investigate whether tumor gene expression patterns could predict clinical response to chemotherapy. Results of these studies are encouraging, indicating that individualization of drug treatment based on multigenic response-predictive markers is feasible .Faced with the skyrocketing expenses of modern healthcare, and microarrays offer a solution with the prospects for more efficient, cost effective, and personalized approaches to patient care.

13.3. 4 MOLECULAR DIAGNOSTIC RESEARCH

Molecular diagnostics is a rapidly advancing field in which insights into disease mechanisms are being elucidated by use of new gene-based biomarkers. Until recently, diagnostic and prognostic assessment of diseased tissues and tumors relied heavily on indirect indicators that permitted only general classifications into broad histologic or morphologic subtypes and did not take into account the alterations in individual gene expression. Global expression analysis using microarrays now allows for simultaneous interrogation of the expression of thousands of genes in a high-throughput fashion and offers unprecedented opportunities to obtain molecular signatures of the state of activity of diseased cells and patient samples. Microarray analysis may provide invaluable information on disease pathology, progression, resistance to treatment, and response to cellular microenvironments and ultimately may lead to improved early diagnosis and innovative therapeutic approaches. Microarray technology has been widely used in the past 3 years to investigate tumor classification, cancer progression, and chemotherapy resistance and sensitivity .

The numbers of microarray based studies identifying new genes or molecular pathways involved in tumor classification, cancer progression, or patient outcome are growing exponentially. We are now in the "postgenomic era", during which the diagnostic, prognostic, and treatment response biomarker genes identified by microarray screening will be interrogated to provide personalized management of patients.

Clinicians will be able to use microarrays during early clinical trials to confirm the mechanisms of action of drugs and to assess drug sensitivity and toxicity. Coupled with more conventional biochemical analysis such as IHC and ELISA, microarrays will be used for diagnostic and prognostic purposes.

Array-based comparative genomic hybridization (CGH) measures copy-number variations at multiple loci simultaneously, providing an important tool for studying cancer and developmental disorders and for developing diagnostic and therapeutic targets(Barrett *et al.,* 2004). The results demonstrate that oligonucleotide arrays designed for CGH provide a robust and precise platform for detecting chromosomal alterations throughout a genome with high sensitivity even when using full-complexity genomic samples.Using cDNA microarrays can acquire gene expression profiles that characterize anticancer drug sensitivity at the same time evaluating the genome-wide gene expression profiles of various cancer cell lines to identify the gastrointestinal tract cancer cell-related genes.

Microarray technology is extensively used in biological research . DNA microarrays are an integral part of the process for therapeutic discovery, optimization and clinical validation. At an early stage, investigators use arrays to prioritize a few genes as potential therapeutic targets on the basis of various criteria. Subsequently, gene expression analysis assists in drug discovery and toxicology by eliminating poor compounds and optimizing the selection of promising leads. Integral to this process is the use of sophisticated statistics, mathematics and bioinformatics to define statistically valid observations and to deduce complex patterns of phenotypes and biological pathways. In short, microarrays are redefining the drug discovery process by providing greater knowledge at each step and by illuminating the complex workings of biological systems .cDNA microarrays have been used to acquire gene expression profiles that characterise anticancer drug sensitivity .With the development of microarray based tests for diagnosing a wide range of dieases ,the ability to gain more information from a limited clinical sample by using highly parallel expression techniques has emerges .Several research groups have focused on identifying subsets of genes that show differential expression (Favis and Barany,2000) between healthy tissues or cell lines and their tumor counterparts to identify biomarkers for several solid tumors, including ovarian carcinomas , oral cancer, melanoma , colorectal cancer , and prostate cancer .SNP analysis raises the possibility of individual tumor genome-wide allelotyping with potential prognostic and diagnostic applications(Hoque *et al.*,2003). They provide a basis for DNA-based cancer classification and help to define the genes being modulated, improving understanding of cancer genesis and potential therapeutic targets (Dutt and Beroukhim ,2007).

Although the major limiting factors for routine use in a clinical setting at present are cost and access to the microarray technology, it is likely that costs will decrease in the near future and that the technology will become increasingly user friendly and automated.

13.4 CONCLUSION

DNA microarray analysis is a fast and versatile approach to acheive high throughput exploration of genome structure , gene expression programm and of gene function at both cellular and organism levels .It is a complex multistep process involving numerous specific equipments and requiring a strong expertise in various areas , including molecular biology ,image analysis , computing , and statistics .

Protein arrays are emerging to follow DNA chips as possible screening tools . In today's proteome era, efforts are undertaken to adapt microarray technology in order to analyse the expression of a large number of proteins simultaneously and screen entire genomes for proteins that interact with particular factors, catalyse particular reactions, act as substrates for protein-modifying enzymes and/or as targets of autoimmune responses . Protein microarray technologies, such as reverse-phase protein arrays, provide the unique opportunity to profile tissues and assess the activity of signalling pathways within isolated cell populations. This technology can be used to identify patients likely to benefit from specific treatment modalities and also to monitor therapeutic response in samples obtained during and after treatment. Routine application of genomic and proteomic microarray technologies in clinical practice will require significant efforts to standardise the techniques, controls and reference standards, and analytical tools used. Extensive, independent validation using large, statistically-powered datasets will also be

necessary. Inclusion of concomitant genomic and proteomic-based molecular profiling techniques into clinical trial protocols will bring us closer to the reality of patient-tailored therapy. Protein microarray technology allows the simultaneous determination of a large variety of parameters from a minute amount of sample within a single experiment. Assay systems based on this technology are currently used for the identification, quantitation and functional analysis of proteins that are of interest for proteomic research in basic and applied biology and for diagnostic applications. Such novel assays are also of major interest for the pharmaceutical industry, focusing on the identification of biomarkers and the validation of potential target molecules. Sensitivity, reproducibility, robustness and automation have to be demonstrated before this technology will be suitable for high-throughput applications. A rapidly expanding use of this technology is the acquisition of information about the posttranslational modifications of proteins reflecting the activity state of signal pathways and networks, and is now employed for the analysis of biopsy samples in clinical trial research.

The examination of gene expression using microarrays holds tremendous promise for the identification of candidate genes involved in a variety of proceses(Tan *et al.,* 2003). The range of applications of microarray technology is enormous. Recent studies in human cancer have demonstrated that microarrays can be used to develop a new molecular taxonomy of cancer, including clustering of cancers according to prognostic groups on the basis of gene expression profiles. The list of potential uses of this technique is not limited to cancer research. For example, the temporal impact on gene expression by drugs, environmental toxins, or oncogenes may be elucidated, and regulatory networks and coexpression patterns can then be deciphered. In the 6 years since its inception, microarray technology has become a major tool for the investigation of global gene expression of all aspects of human disease and in biomedical research.

Predicting the concerning the future of the gene chip market has been all over the map. Only a few years ago, the market was described as exploding and surging . Indeed the sky seemed to be the limit. Drug discovery had a virtually unlimited appetite for the enabling technologies and the possibilities of expanding chip technology beyond DNA chips to protein, tissue and cell microarrays and eventually lab-on chip technologies looked like a road without end, one that would lead directly into clinical diagnostics and promised land of personalised medicine. In the future, technologies such as gene chip array will enhance genetic medicine and provide novel insights into a patient's susceptibility to disease, enabling a better assessment of prognostic risk factors, quicker diagnosis, and accurate prediction of individual responsiveness to drugs. The predictable consequences of such an approach on the prevention and treatment of diseases could revolutionize medicine. Microarrays are on their way to revolutionize and to allow us to understand how life really works.

REFERENCES

1. Barrett, J.C., Kawasaki, E.S., 2003, Microarrays: the use of oligonucleotides and cDNA for the analysis of gene expression. *Drug Discov Today* 8:134–141.
2. Brenton, J.D., Aparicio, S.A., Caldas, C., 2001 Molecular profiling of breast cancer: portraits but not physiognomy. *Breast Cancer Res* 3:77–80.
3. Cargill, M., Altshuler, D., Ireland, J., Sklar, P., Ardlie, K., Patil, N., Shaw, N., Lane, C.R., Lim, E.P., Kalyanaraman, N., Nemesh, J., Ziaugra, L., Friedland, L., Rolfe, A., Warrington, J., Lipshutz, R., Daley, G.Q.,

Lander, E.S., 1999, Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nature Genet* 22: 231–238.
4. Culhane, A.C., Perriere, G., Higgins, D.G., 2003, Cross-platform comparison and visualisation of gene expression data using co-inertia analysis. *BMC Bioinformatics* 4:59.
5. Dutt, A., Beroukhim, R., 2007, Single nucleotide polymorphism array analysis of cancer. *Curr Opin Oncol* 19:43-49.
6. Favis, R., Day, J.P., Gerry, N.P., Phelan, C., Narod, S., Barany, F., 2000, Universal DNA array detection of small insertions and deletions in BRCA1 and BRCA2. *Nat Biotechnol* 18:561–564.
7. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S., 1999, Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531–537.
8. Hacia, JG., 1999, Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genet* 21: 42–47.
9. Hacia, J.G., Brody, L.C., Chee, M.S., Fodor, S.P., Collins, F.S., 1996, Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-color fluorescence analysis. *Nature Genet* 14:441–447.
10. Hoque, M.O., Lee, C.C., Cairns, P., Schoenberg, M., Sidransky, D., 2003, Genome-wide genetic characterization of bladder cancer: a comparison of high-density single-nucleotide polymorphism arrays and PCR-based microsatellite analysis. *Cancer Res* 63:2216-2222.
11. Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., Linsley, P.S, 2001, Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19: 342–347.
12. Johnson, R.A., Wichern, D.W., 2002, Applied Multivariate Statistical Analysis, 5th edn. Prentice Hall, Upper Saddle River, NJ.
13. Kane, M.D., Jatkoe, T.A., Stumpf, C.R., Lu. J., Thomas, J.D., Madore, S.J., 2000, Assessment of the sensitivity and specificity of oligonucleotide (50mer) microarrays. *Nucleic Acids Res* 28:4552–4557.
14. Kothapalli, R., Yoder, S.J., Mane, S., 2002, Microarray results: how accurate are they? *BMC Bioinformatics* 3: 22.
15. Kuo, W.P., Jenssen, T.K., Butte, A.J., Ohno-Machado, L., Kohane, I.S., 2002, Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 18: 405–412.
16. Li, C., Hung Wong, W., 2001, Model-based analysis of oligonucleotide arrays: model validation, design issues and standard error application. *Genome Biol* 2:research/0032.1-0032.11.
17. Lipshutz, R.J., Fodor, S.P, Gingeras, T.R., Lockhart, D.J., 1999, High density synthetic oligonucleotide arrays. *Nature Genet* 21:20–24.
18. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., Brown, E.L., 1996, Expression monitoring by ybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675–1680.
19. Mecham, B.H., Klus, G.T., Strovel, J., Augustus, M., Byrne, D., Bozso, P., Wetmore, D.Z., Mariani, T.J., Kohane, I.S., Szallasi, Z., 2004, Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. *Nucleic Acids Res* 32: e74.
20. Schmoock, G., Ehricht, R., Melzer, F., Rassbach, A., Scholz, H.C., Neubauer, H., Sachse, K., Mota, R.A., Saqib, M., Elschner, M., 2009, DNA microarray-based detection and identification of Burkholderia mallei, Burkholderia pseudomallei and Burkholderia spp. *Mol Cell Probes* PMID: 19366627.
21. Steven, A., Greenberg, M.D., 2001, DNA microarray gene expression analysis technology and its application to neurological disorders. *Neurology* 57:755-761.
22. Stoevesandt, O., Taussig, M.J., He, M., 2009, Protein microarrays: high-throughput tools for proteomics. *Expert Rev Proteomics* 6: 145-157.
23. Tan, P.K., Downey, T.J., Spitznagel, E.L, Jr, Xu, P., Fu, D., Dimitrov, D.S., Lempicki, R.A., Raaka, B.M., Cam, M.C., 2003, Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 31: 5676–5684.
24. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., Church, G.M., 1999, Systematic determination of genetic network architecture. *Nature Genet* 22: 281–285.

25. Wang, H,Y., Malek, R.L., Kwitek, A.E., Greene, A.S., Luu, T.V., Behbahani, B., Frank, B., Quackenbush, J., Lee, N.H., 2003, Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. *Genome Biol* 4:R5.
26. Wright, G., Tan, B., Rosenwald, A., Hurt, E.H., Wiestner, A., Staudt, L.M., 2003, A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci* USA, 100:9991–9996.
27. Yuen, T., Wurmbach, E., Pfeffer, R.L., Ebersole, B.J., Sealfon, S.C., 2002, Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res* 30:e48.