Genome Resource Facility

Microarray Overview

- Introduction
- What is DNA microarray technology
- Types of DNA microarray technology
- Glass cDNA Microarrays
- Advantages of cDNA Microarrays
- Disadvantages of cDNA Microarrays
- *In situ* oligonucleotide array format
- Advantages of *in situ* oligonucleotide array format
- Disadvantages of *in situ* oligonucleotide array format
- Principles of DNA microarray experiments
- Application

Introduction

DNA microarray technology (also known as DNA arrays, DNA chips or biochips) represents one of the latest breakthroughs and indeed major achievements in experimental molecular biology. This novel technology, which started to appear during the second half of the 1990s, has historically evolved from the initial experimental reports published in the mid 1970s which indicated that labelled nucleic acids could be used to monitor the expression of nucleic acid molecules attached to a solid support. However, it was not until 1995 that the first article describing the application of DNA microarray technology to expression analysis was published in the scientific literature by Patrick Brown and his colleagues at Stanford University (Brown., *et al*. 1995).

Today, however, there is prominent evidence that the technology has made a dramatic advancements since its development and gained an increasing popularity among scientific researchers. It is also unquestionable that many scientific researchers are presumptive about the novelty of this technology regarding it as an indispensable as well as a needful research tool.

Such widespread adoption of DNA microarray technology in both industry and many academic research laboratories, was largely due to its aptitude to provide scientific researchers the opportunity to quickly and accurately perform simultaneous analysis of literally thousands of genes in a massively parallel manner, or even entire genome of an organism e.g. (Bacteria, Yeast, Virus, Protozoa, Mouse or Human) in a single experiment, hence providing extensive and valuable information on gene interaction and function.

Here, our aim is to give a brief overview of cDNA microarray technology, particularly explaining (Watson., *et al*. 1998) what is a cDNA microarray technology (Sinclair., 1999) the various types/platforms of cDNA microarray technology, highlighting both their advantages and disadvantages as well as how they are fabricated or manufactured (Burgess., 2001) the technology's basic fundamental principles and how it works are also outlined, as well as potential applications.

However, for detailed information of the potential and the scientific value of cDNA microarray technology in modern research, the reader is advised to look up the large number of excellent reviews of DNA array technology available (Rhodius., *et al* 2002). We particularly recommend the January 1999 supplement of Nature Genetics (Bowtell., 1999).

What is DNA microarray technology?

In it's broadest term, DNA microarray technology may be defined as a high-throughput and versatile technology used for parallel gene expression analysis for thousands of genes of known and unknown function, or DNA homology analysis for detecting polymorphisms and mutations in both prokaryotic and eukaryotic genomic DNA.

However, in its precise and accurate definition DNA microarray is an orderly arrangement of thousands of identified sequenced genes printed on an impermeable solid support, usually glass, silicon chips or nylon membrane.

Each identified sequenced gene on the glass, silicon chips or nylon membrane corresponds to a fragment of genomic DNA, cDNAs, PCR products or chemically synthesised oligonucleotides of up to 70mers and represents a single gene.

Usually a single DNA microarray slide/chip may contain thousands of spots each representing a single gene and collectively the entire genome of an organism. A schematic diagram of the two most commonly used DNA microarray formats to date are shown in Figure 1 and 2.

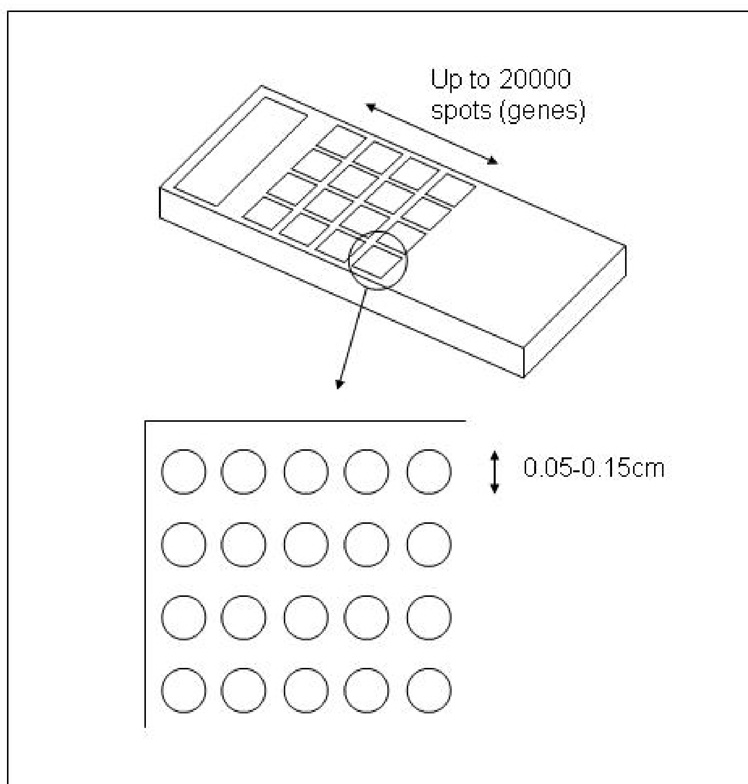

Figure 1. Glass complementary DNA (cDNA) microarray produced by using high-speed precision robot. This type of DNA microarray can bear between 10,000 - 20,000 spots (genes) on an area of 3.6 cm2. Each spot represents the product of a specific gene and is generated by depositing a few nano liters of PCR product representing that specific gene usually at concentration of 100-500 µg/ml. The diameter of each spot is also typically 50-150 µm.

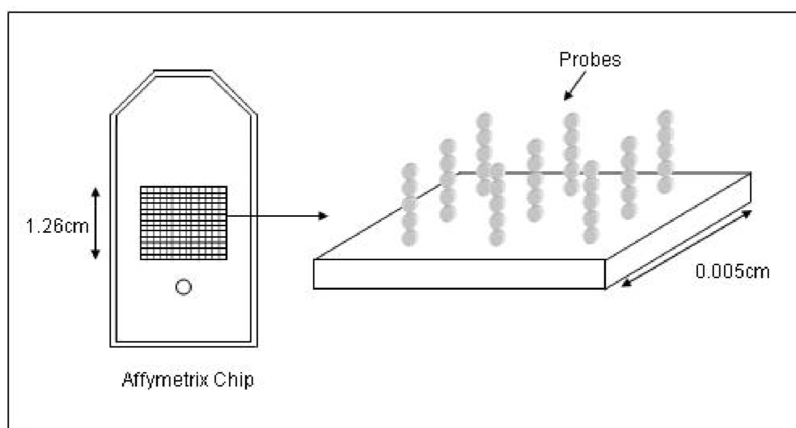

Figure 2. Illustration of a DNA GeneChip (Affymetrix).

Types of DNA microarrays

There are currently two platforms/types of DNA microarrays that are commercially available.

1. Glass DNA microarrays which involves the micro spotting of pre-fabricated cDNA fragments on a glass slide.
2. High-density oligonucleotide microarrays often referred to as a "chip" which involves *in situ* oligonucleotide synthesis.

However, from a manufacturing point of view, there are fundamental differences between the two platforms in regard to the sizes of printed DNA fragments, the methods of printing the DNA spots on the slide/chip, and also the data images generated.

Glass cDNA microarrays

Glass DNA microarrays was the first type of DNA microarray technology developed. It was pioneered by Patrick Brown and his colleagues at Stanford University and is produced by using a robotic device, which deposits (spots) a nanoliter of DNA (50-150 µm in diameter) onto a coated microscope glass slide surface in serial order with a distance of approximately 200-250 µm from each other, one spot-one gene. These moderate sized glass cDNA microarrays also bear about 10,000 spots or more on an area of 3.6 cm2.

As the name suggests, glass cDNA microarrays use specially manufactured glass slides with desired physico-chemical characteristics e.g. excellent chemical resistance against solvents, good mechanical stability (increased thermal strain point) and low intrinsic fluorescence properties.

However, to produce a complete whole genome glass DNA microarray, a series of consecutive steps are followed, ideally each step requiring an appropriate and careful approach. Here, we will not discuss in detail how each step is performed, but briefly outline these steps in the order they are followed.

The first step of manufacturing a glass cDNA microarray is selecting the material to spot onto the microscope glass surface e.g. the genes from public databases/repositories or institutional sources. This is followed by the preparation and purification of DNA sequences representing the gene of interest. In the preparation process, PCR is used to amplify the DNA from library of interest using a universal primers or gene specific primers and the purity of the DNA fragments representing genes of interest are generally checked by sequencing or using on agarose gel to concomitantly obtain an estimate of the DNA concentration. This is an important step because all the DNA fragments should be of similar concentration/molarity and size, to achieve similar reaction kinetics for all hybridisations. The third step is spotting DNA solution onto chemically modified glass slides usually with poly(L-lysine) or other cross-linking chemical coating materials such as polyethyleneimine polymer p-aminophenyl trimethoxysilane/diazotization chemistry and dendrimeric structure. It is these substrates that are coated on the surface of the glass slide that determines how the DNA solution will be immobilised on the surface e.g. covalent or non covalent. However in the course of poly(L-lysine) the negatively charged phosphate groups in the DNA molecule, form an ionic bond with the positively charged amine-derivatised surface. This spotting step is achieved via a contact printing using precisely controlled robotic pins or other equivalent delivering technology such as inkjet printing.

The last step of manufacturing glass DNA microarrays is the post-print processing step involving the drying of the DNA on the slide overnight at room temperature and the use of UV cross-linking to prevent subsequent binding of DNA, and to decrease the background signal upon hybridisation of a labelled target.

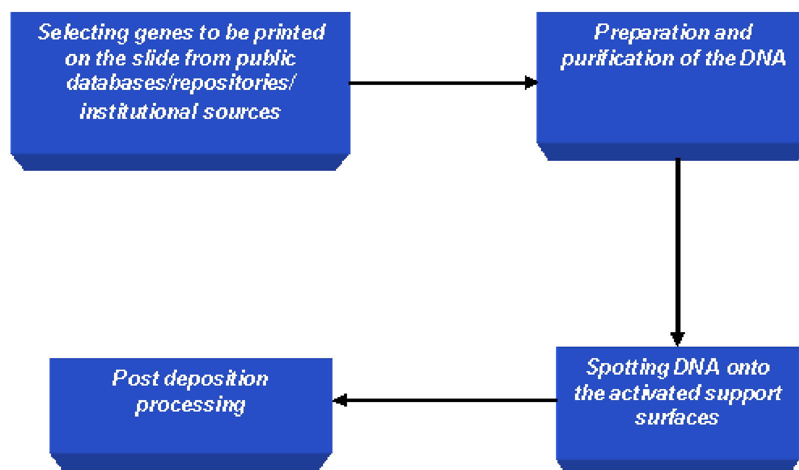

Figure 3. Steps of manufacturing glass cDNA microarrays.

Advantages of cDNA microarrays

Advantages of Glass cDNA microarrays include their relative affordability with a lower cost. Its accessibility requiring no specific equipment for use such that hybridisation does not need specialised equipment, and data capture can be carried out using equipment that is very often already available in the laboratory and flexibility of design as necessitated by the scientific goals of the experiment. In addition to that, Glass cDNA microarrays also have increased detection sensitivity due to longer target sequences (2 kbp).

Disadvantages of cDNA microarrays

Despite their wide spread use, glass cDNA microarray have a few disadvantages such as intensive labour requirement for synthesising, purifying, and storing DNA solutions before microarray fabrication. Further, more printing devices required thus making microarrays more expensive. Also during microarray experiments in the laboratory, sequence homologies between clones representing different closely related members of the same gene family may result in a failure to specifically detect individual genes and instead may hybridise to spot(s) designed to detect transcript from a different gene. This phenomena is known as cross hybridisation.

in situIn situ oligonucleotide array format

In situ (on chip) oligonucleotide array format is a sophisticated platform of microarray technology which is manufactured by using the technology of *in situ* chemical synthesis that was first developed by Stephen Fodor *et al*. (1991). However, the industry leader in the field of *in situ* oligonucleotide microarrays (Affymetrix) has further pioneered this type of technology to manufacture so-called GeneChips which refers to its high density oligonucleotide based DNA arrays.

Presently, the commercial versions of Affymetrix GeneChips hold up to 500,000 probes/sites in a 1.28-cm2 chip area, and due to such very high information content (genes), they are finding widespread use in the hybridisation-based detection and analysis of mutations and polymorphisms, such as single nucleotide polymorphisms or disease-relevant mutations analysis ("genotyping"), as well as a wide range of other applications such gene expression studies, to mention a few.

The basic principles of manufacturing Affymetrix's GeneChips is the use of photolithography and combinatorial chemistry to manufacture short single strands of DNA onto 5-inch square quartz wafers. Unlike glass cDNA, the genes on the chip are designed based on sequence information alone, and then using an industry chip synthesiser, sequences are directly synthesised onto the surface of the 5-inch square quartz wafer at a pre-selected positions.

Detailed stepwise synthesis of *in situ* synthesis of oligonucleotides (Affymetrix GeneChips) is beyond the scope of this overview, but we will briefly outline some of the concepts related to the fabrication process.

The fabrication process of Affymetrix's GeneChips using a DNA photolithography process starts by the derivatization of the solid support, usually quartz with a covalent linker molecule terminated with a photolabile protecting group. This is firstly achieved by washing the quartz to ensure uniform hydroxylation across its surface and then placing it in a bath of silane, which reacts with the hydroxyl groups of the quartz and forms a matrix of covalently linked molecules.

The *in situ* synthesis of oligonucleotides occur in parallel, resulting in consecutive addition of A, C, G and T nucleotides to the appropriate gene sequences on the array. At each step in the synthesis process, oligonucleotide chains that for example require adenine in the next position are deprotected by light at the appropriate positions by a mask. The quartz (chip) is then flooded with a solution containing activated adenine nucleotides with a removable protection group, which are coupled to the deprotected positions. Uncoupled adenine residues are washed away and another mask is applied to further carry out the deprotection of the next nucleotide. Finally, repeating the process ~70 times, with 70 different masks, allows synthesis of the complete array of thousands of 25-mer oligonucleotides in parallel.

Advantages of *in situ* oligonucleotide array format

Advantages offered by the *in situ* oligonucleotide array format include speed, specificity and reproducibility. Speed, in terms of generating the array is prime advantage because, spotting the DNA onto the chip requires only that the DNA sequence of interest be known, therefore no time is spent in the handling of cDNA resources such as the preparation and accurate determination of handling bacterial clones, PCR products, or cDNAs, thus reducing the likelihood of contamination and mix up. However, before manufacturing the array, prior knowledge of the genome sequence is required to design the oligonucleotide sets, and when this is not available, alternative methods of printing isolated genetic material may be preferred.

Other advantages of the *in situ* oligonucleotide array format include high specificity and reproducibility. Both of these attributes are due to the way oligonucleotide sequences to be printed on the chip are designed and the use of multiple, short sequence(s) representing the unique sequence of genes. For example, when designing oligonucleotide sequences for a gene, each sequence is designed to be perfectly complementary to a target gene sequence, at the same time an additional partner sequence is designed that is identical except for a single base mismatch in its centre. This sequence mismatch strategy, along with the use of multiple sequence(s) for each gene increases specificity and helps to identify and minimise the effects of non-specific hybridisation and background signal. This strategy also allows the direct subtraction of cross-hybridisation signals and discrimination between real and non-specific signals.

Disadvantages of *in situ* oligonucleotide array format

There are several disadvantages to the *in situ* oligonucleotide array format including practical limitations in terms of affordability and flexibility. Firstly, *in situ* oligonucleotide array formats tend to have expensive specialised equipments e.g. to carry out the hybridisation, staining of label, washing, and quantitation process. Secondly, ready made in situ oligonucleotide array format (GeneChips) are still expensive, although there has been reductions in cost as the market of microarrays has expanded. Thirdly, although short-sequences used on the array confer high specificity, they may have decreased sensitivity/binding compared with glass cDNA microarrays. Such low sensitivity however is compensated for by using multiple probes.

In situ oligonucleotide array format also offers reduced flexibility although this is not the case with respect to the array design. However, there are occasions when the production of the array, hybridisation and detection equipment are restricted to centralised manufacturer facilities, thus limiting the researcher's flexibility. Similarly, the cost and time needed to manufacture the *in situ* oligonucleotide array format makes it uneconomical for an average laboratory to synthesise its own chips.

Principles of DNA Microarray experiments

The principle of DNA microarray technology is based on the fact that complementary sequences of DNA can be used to hybridise immobilised DNA molecules. This involves three major multi-stage steps;

1- Manufacturing of microarrays: This step involves the availability of a chip or a glass slide with its special surface chemistry, the robotics used for producing microarrays by spotting the DNA (targets) onto the chip or for their *in situ* synthesis.
2- Sample preparation and array hybridisation step: This step involves mRNA or DNA isolation followed by fluorescent labelling of cDNA probes and hybridisation of the sample to the immobilised target DNA.
3- Image acquisition and data analysis: Finally, this step involves microarray scanning, and image analysis using sophisticated software programs that allows us to quantify and interpret the data.

However, here, we will concentrate on how microarray experiments are performed in the laboratory, rather than the technological developments involving array construction or manufacturing using precision robotic devices.

Typically, a microarray experiment involves the comparison of a query or experimental sample representing the expression pattern of genes in a specific set of conditions, with a control sample representing all the genes that are expressed in the cells/tissue to be analysed.

An example of this is comparisons made between expression profiles of bacteria within infected cells (query) and the same bacteria cultured under standardised *in vitro* conditions of growth (control). Or similarly a comparison made between an isogenic mutant and the wildtype strain.

There are four major steps in performing a typical microarray experiment.

1. Sample preparation and labelling
2. Hybridisation
3. Washing
4. Image acquisition and Data analysis

Sample preparation and labelling

There are a number of different ways in which a DNA microarray sample is prepared and labelled. All of these different approaches however have their own advantages and disadvantages with respect to many factors such as the starting amount of RNA or DNA required, through to cost, time and data acquisition and transformation. However the choice of which one to use depends on these factors as well as the type of microarray technology used, for example the slide type and the detection equipment. Here, we used the term "sample" to refer to the free, fluorescently labelled cDNA, not to confuse with the immobilised DNA known as reporter element (s)

Initially, the sample preparation starts by isolating a total RNA containing messenger RNA that ideally represents a quantitative copy of genes expressed at the time of sample collection (experimental sample & reference sample). This step is crucial, simply because the overall success of any microarray experiment depends on the quality of the RNA.

For example purity in the sense of homogeneity or uniformity of the mRNA is a critical factor in the downstream hybridisation performance, particularly when fluorescence is used, as cellular proteins, lipids, and carbohydrates can mediate significant nonspecific binding of labeled cDNAs to matrix surfaces. The sample mRNA extracted from the biological sample of interest and the reference are then separately converted into complementary DNA (cDNA) using a reverse-transcriptase enzyme. This step also requires a short primer to initiate cDNA synthesis. Next, each cDNA (Sample and Control) are labelled with a different tracking molecule, often fluorescent cyanine dyes (i.e. Cy3 and Cy5)

Array hybridisation

Hybridisation is the process of joining two complementary strands of DNA to form a double-stranded molecule. Here, the labelled cDNA (Sample and Control) are mixed together, and then purified to remove contaminants such as primers, unincorporated nucleotides, cellular proteins, lipids, and carbohydrates. Purification is usually carried out using filter spin columns such as Qiaquick from Qiagen. After purification, the mixed labelled cDNA is competitively hybridised against denatured PCR product or cDNA molecules spotted on a glass slide. Ideally, each molecule in the labelled cDNA will only bind to its appropriate complementary target sequence on the immobilised array.

Before hybridisation however, the microarray slides are incubated at high temperature with solutions

of saline-sodium buffer (SSC), Sodium Dodecyl Sulfate (SDS) and bovine serum albumin (BSA) to reduce background due to nonspecific binding.

The slides are washed after hybridisation, first to remove any labelled cDNA that did not hybridise on the array, and secondly to increase stringency of the experiment to reduce cross hybridisation. The later is achieved by either increasing the temperature or lowering the ionic strength of the buffers.

Image acquisition and data analysis is the final step of microarray experiments. The aim is to produce an image of the surface of the hybridised array. Here the slide is dried and placed into a laser scanner to determine how much labelled cDNA (probe) is bound to each target spot. Laser excitation of the incorporated targets yield an emission with characteristic spectra, which is measured using a confocal laser microscope. Classically, microarray software often uses green spots on the microarray to represent genes upregulated compared to control, red to represent those genes that are downregulated in the experimental sample, and yellow to represent those genes of equal abundance in both experimental and control samples.

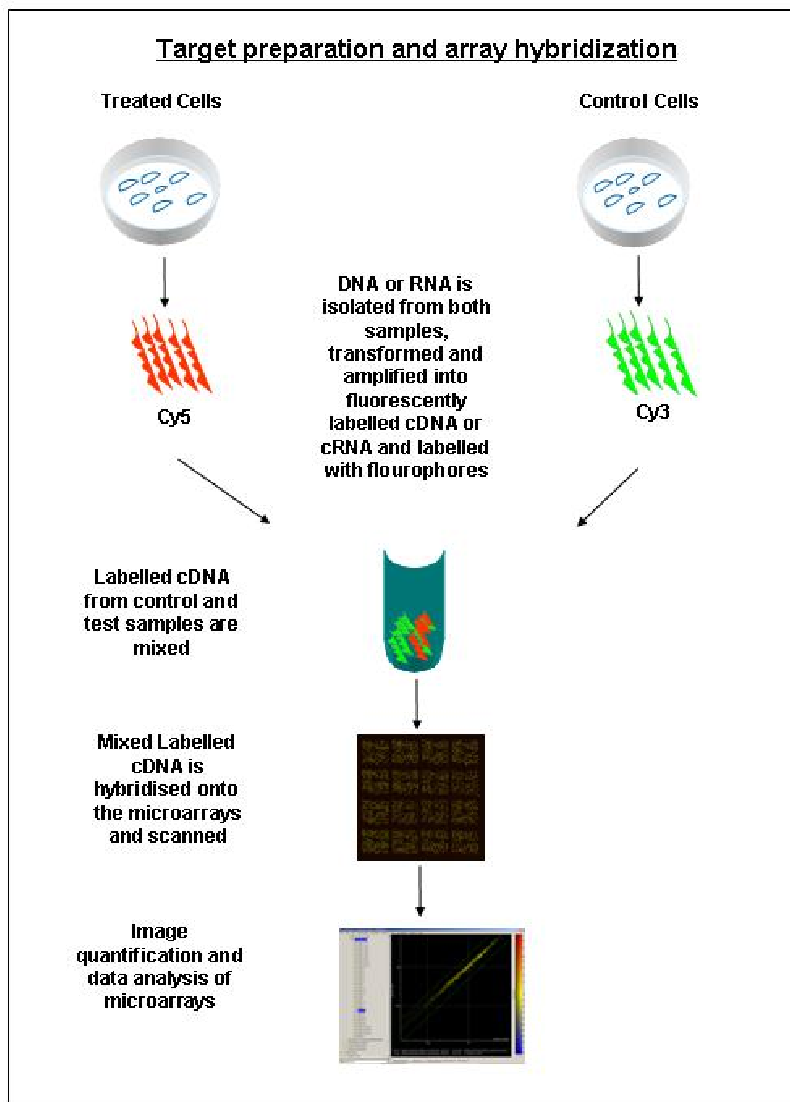

Figure 4. Microarray experimental principles.

Applications of DNA Microarray Technology

The range of applications for microarray technology is enormous. However, although in depth details of each one is beyond the scope of this overview, there are two distinctive applications of microarrays that are in wide spread use.

1. Gene expression profiling to measure the expression of genes between different cell populations.
2. Comparative genomics to analyse genomic alterations such as sequence and single nucleotide polymorphisms.

Microarray as a gene expression profiling tool

The principle aim of using microarray technology as a gene expression profiling tool is to answer some of the fundamental questions in biology such as "when, where, and to what magnitude genes of interest are expressed.'' Clearly, if a gene is not expressed in a defined time/condition at a defined cell compartment, then it is possible to imagine it can play no role in that subnetwork. This approach is based on the assumption that cellular genes are expressed in response to a particular state and that the expression profile represents the subset of gene transcripts or mRNA expressed in a cell or tissue. In addition to that, expression profiling by microarray analysis provides another approach to measure changes in the multigene patterns of expression to better understand about regulatory mechanisms and broader bioactivity functions of genes.

It is therefore appreciable that the knowledge obtained from microarray gene expression analysis will probably increase our basic understanding of the cause and consequences of diseases (pathogenesis), how drugs and drug candidates work in cells and organisms, and what gene products might have therapeutic uses or may be studied further as an appropriate targets for therapeutic intervention.

For example, in the context of microbiology, microarray gene expression is used to analyse complex cellular behaviour and to explore the complex interaction between host and microbial pathogens. This ambitious and plausible attempt to understand such molecular interaction is achieved by directly comparing gene expression profiles of host cell to the expression profile of the pathogen. Also an *ex vivo* measurement of gene expression for host cells before and after they are infected with a microbial pathogen can greatly increase our understanding of such complex molecular interplay. Furthermore by following the pattern of gene expression at different times, it is possible to elucidate which host or pathogen genes are up or downregulated over the course of infection to further identify critical target genes and drug-specific targets in both host and microbial pathogens.

Microarray as a comparative genomics tool

Another important application of microarray technology, which is also finding a widespread use is gene mutation analysis to analyse genomic alterations such as sequence and single nucleotide polymorphisms. Currently, in the context of microbiology microarray gene mutation analysis is directed to characterisation of genetic differences among microbial isolates, particularly closely related species. For example detecting the presence or absence of DNA sequences/gene(s) between pathogenic and non-pathogenic strains of the same species. Such informative approach would allow us to reveal genes exclusively present in the former that may be required for infectivity, virulence or adaptation to a particular host niche

Similarly, DNA microarray technology can be used to compare between a fully sequenced genome and an unsequenced but related genome of closely related bacteria. Interestingly, such approach can provide us a valuable information about the diversity and evolution of pathogens and symbionts. Comparisons of this kind use a microarray containing representations of all the open reading frames (ORFs) of the sequenced, reference strain and labelled DNA from the unsequenced, experimental strain. The resulting hybridised array will then reveal genes common to both strains and genes that are present in the reference strain but absent in the experimental strain. This method, however, may not detect genes present in the experimental strain, but missing in the reference strain, although the use of multi-genome (species array) may detect genes present in the experimental strain. This method may not also detect point mutations, including frame-shift mutations, small deletions and deletions in homologous repetitive elements, rearrangements of the genome that have not resulted in deletion of a gene from the experimental strain. However the use of affymetrix gene chip can identify these mutations.

An elegant and well known example of using microarray as a comparative genomics tool is the comparison made by Behr et al., 1999 between several *Mycobacterium bovis* vaccine strains e.g. the genome composition of the sequenced M. tuberculosis laboratory strain H37Rv with the closely related pathogenic species, M. bovis, and with several strains of the bacille Calmette-Guerin (BCG) vaccine variant that was produced by serial *in vitro* passage of M. bovis between 1908 and 1921.

This particular study, has revealed several chromosomal deletions in the different vaccine strains in comparison to their progenitor, these deletions are possibly thought to be responsible for the variable effectiveness of the BCG vaccine seen worldwide.