DNA MICROARRAYS: A HIGH-THROUGHPUT TECHNOLOGY

Nathalie Corneau, Franco Pagotto and Jeffrey M. Farber, Health Canada, Ottawa, Ontario. Phone: (613) 952-4577; Email: nathalie_corneau@hc-sc.gc.ca

In the last few years, DNA microarray technology has emerged as one of the most powerful tools in the biotechnology field. Microarrays (also known as biochips or chips) are a series of DNA molecules of known sequences (called probes), fixed on a substrate (glass or nylon) at a defined location. Microscope glass slides are the most commonly used support for high density DNA microarrays. Although standard microscope slides may be used, they must be treated to increase the adherence of DNA to the glass. Researchers can purchase ready to print slides or a variety of ready to use microarrays. Many biotechnology companies now sell mouse, rat, yeast, bacteria and human DNA chips. The DNA molecules used to develop DNA arrays (*i.e.*, to print on the slide) can be partial gene sequences generated by polymerase chain reaction (PCR), full-length cDNAs (DNA molecule generated form a mRNA molecule using the Reverse Transcriptase (RT) enzyme) or oligonucleotides. The number of spots deposited on a slide range from a few hundred to many thousands.

Even though DNA microarrays are primarily used to study genomics, the technology is still very expensive. To set up a DNA microarray facility, a laboratory requires an investment of at least a quarter million Canadian dollars. The most important equipment is a two color, confocal scanner ($60 000 - $100 000) with a computer powerful enough to manage the data ($3 000). A good array software for data acquisition and image analysis is also required ($5 000). Laboratories that intend to print their own chips will need an arrayer ($80 000 - $140 000) and blank slides ($5 - $10 per slide). On the other hand, one may choose to use a commercially available DNA chip. The cost for these range from $300 and $600 per slide, depending on what material is printed on the chip. Companies now offer automated stations ($85 000) to facilitate reproducibility of the hybridization steps where the target DNA is annealing with the probe.

DNA microarray technology allows the researcher to screen either selected genes or an entire genome at once. Applications for chip technology include gene discovery and expression, detection of mutations or polymorphisms, sequencing, as well as detection and molecular typing of pathogens. The amount of data obtained from a single experiment is huge and must be carefully analyzed using sophisticated software(s).

In most applications, DNA arrays are used as a tool for examining differential gene expression using two fluorescent dyes (Cy3 and Cy5) in which both experimental and control samples are hybridized to the same array (Fig.1). In this example, targets (*i.e.*, free nucleic acid samples whose identity and/or abundance are being detected) have been prepared from mRNA expressed in cells grown in aerobic or anaerobic conditions. Studies based on bacterial organisms are more complicated. Reproducible purification of bacterial mRNA is very difficult because of its low concentration in total RNA. In addition, bacterial mRNA degrades easily and does not contain a poly-A tail (found in eucaryote mRNA) to facilitate its purification. In these instances, standardised protocols must be developed for a successful experiment, as different RNA extraction methods can give different results. There are three ways to prepare targets from RNA; 1) RNA can be labelled directly with the fluorophores Cy3 or Cy5 (most commonly used dyes); 2) a cDNA is produced and the Cy3 or Cy5 conjugated nucleotides are incorporated during the reverse transcriptase (RT) reaction; 3) the dyes can be incorporated in a PCR assay (following the initial RT-

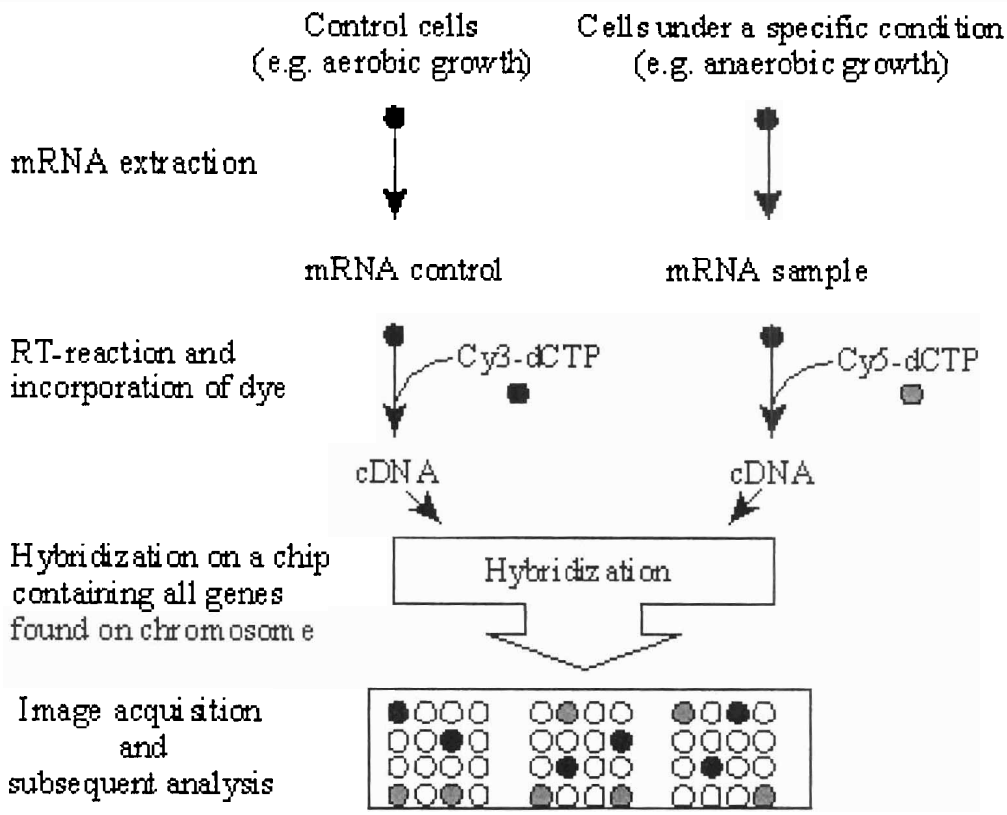

Figure 1. Illustration of differential gene expression analysis under aerobic and anaerobic growth. Different dyes allow the comparison between RNA samples from the same bacterium grown under two entirely different conditions in the same assay. ● More expression in the control (aerobic) sample (*i.e.*, same gene is repressed or turned off in the anaerobic sample); ● Greater expression of genes under anaerobic growth (as compared to aerobic growth); ○ Gene expression is similar in both aerobic and anaerobic samples.

reaction) to increase target sensitivity. In other applications, targets can be made from chromosomal DNA or even antibodies (*e.g.*, in protein arrays).

In array probe-target detection experiments, hybridization and washing are critical steps in generating high quality data. As glass is a non-porous substrate, the hybridization volume can be kept to a minimum (8-100 uL), thus enhancing the kinetics of annealing targets to probes. The classical method of hybridization is usually done under a cover-slip. That is, the labelled RNA is pipetted on the printed slide, and a cover-slip is placed over the array(s). This setup allows for random and passive interactions between the probes and their target(s) to occur. Hybridization cassettes have been developed to maintain a saturated humidity, thereby avoiding evaporation of the targets. An alternative to this method is to use a specially designed cassette and cover the whole chip with the hybridization solution (~ 100 uL). This method is gaining in popularity since the circulation of fluid can be controlled to allow a better interaction between DNA molecules (*i.e.*, probe and target). Like a standard Northern experiment, for the detection of RNA molecules of interest, temperature and salt concentration must be adjusted for optimal hybridization to occur. Temperature is con-

trolled by a heated plate beneath the slide or, alternatively, hybridization cassettes can be used in water baths. Stringency can be modified by temperature variation, the use of different salt concentrations, or the addition of sodium dodecyl sulfate (SDS) in the washing solutions in order to optimize the hybridization process. Finally, the slides are dried, usually by centrifugation, before analysis.

In an array-based assay, sensitive and accurate signal output detection is imperative. A two-color confocal scanner is used for generating light of the correct excitation wavelength (550 nm for Cy3; 650 nm for Cy5) for detection. The resulting image is easily imported into most analysis software. In theory, a DNA microarray is made of perfectly aligned spots with the same diameter, in a uniform background; in reality, this rarely occurs. This is in part the reason why specialized programs are needed to analyse the images obtained.

There are many steps required in the analysis of microarrays. The first step in data analysis is the localisation of each spot. Most programs are flexible enough to capture and analyse a spot located anywhere on a chip. The software has complex algorithms that take into account the imperfections in the gridding process and the non-uniform background. To complete the analysis and allow the comparison of each positive hybridization signal, the background noise corresponding to the area of each spot has to be subtracted from the positive result. Following this step, the pixel intensity of each spot is calculated to be proportional to the number of dye-labelled cDNA hybridized with the immobilized molecules. Quantitation of gene expression is then possible by the analysis of a variety of known controls present on the same chip.

DNA microarray technology is currently viewed as a powerful tool for the study of genomics. In fact, the benefits of microarray technology are much greater than this. It is particularly well suited to detect large numbers of probe sequences. Within the area of food biotechnology, DNA chips can be used for analysing transcription profiles of foodborne pathogens, as well as useful bacteria such as the lactic acid bacteria. The data obtained can provide information on gene expression which is useful for studying pathogenicity, stress conditions responses, antimicrobial resistance, metabolic pathways, mutations and phage resistance. What makes microarray technology appealing is its applicability to a wide variety of apparently non-related research interests.

Both fundamental and applied research can benefit from this technology. For example, dairy food industries can develop new cocktails of microorganisms for cheese fermentation, reducing the fermentation failure due to phage infection and characterizing genes responsible for organoleptic properties. On the other hand, academic institutions are focussing their efforts on fundamental research such as metabolic gene expression, cancer characterization and antibiotic resistance. Government laboratories can enhance their regulatory roles by developing DNA arrays for the detection of genetically modified foods and/or organisms, as well as foodborne pathogen detection and identification.

The first eucaryotic chip was developed for *Saccharomyces cerevisiae* (Lashkari *et al.*, 1997). Sequences representing the entire genome were spotted as a grid onto a single microscope slide (*i.e.*, up to 2 479 open reading frames). Following this publication, the use of microarray technology for the food industry has been described. Of the latter, the most studied has been the genetics of cereals such as rice, corn and soybean (McGonigle *et al.*, 2000; Yazaki *et al.*, 2000). Microarrays have also been developed for the comprehensive analysis of plant diseases (Schenk *et al.*, 2000). Major foodborne pathogens (*i.e.*, *Listeria*, *Campylobacter* and *Escherichia coli*) have al-

ready had their partial or complete genomes printed on various chips. Moreover, biotechnology companies now offer ready-to-print oligonucleotide sets representative of the open reading frames for the entire genome of many eucaryotic and procaryotic organisms.

A specific example of the use of DNA microarray technology in the area of food safety is the recent development of a *Listeria* array. In Canada, an average of 50 human listeriosis cases are reported each year (Farber and Peterkin, 2000). Because the majority of listeriosis cases are caused by only three serotypes (1/2a, 1/2b and 4b), improved molecular typing methods are constantly being developed to differentiate isolates. A DNA microarray has been developed in our laboratory for the detection and typing of *L. monocytogenes* from food, clinical, or environmental samples. The expression patterns of different strains is obtained by the hybridization of targets (cDNA molecules representative of total cell's RNA) to the chip. Each expression pattern is analysed and then compared to the other. This method of identification is very sensitive and could be used as a diagnostic platform for improving food safety. Moreover, the specific design of this chip allows the virulence potential determination of various *L. monocytogenes* isolates to be assessed. The use of animal models may eventually be replaced by *in vitro*, microarray-based, pathogenicity assays.

Microarray technology is currently having its biggest impact on the understanding of gene function and expression. It is being used to complement and/or replace traditional, labour-intensive methods, such as Northern analysis. The actual limits of the chips are directly linked to the researcher's imagination. With the evolution of this technology, more economical ways to make and analyse arrays will be developed. Microarrays can benefit from fundamental studies on the technology itself, *i. e.*, DNA interaction, hybrid stability, probe attachment, etc. The understanding of such basic biophysical and biochemical interactions will greatly speed up technological developments in the field and make its use in all areas of science even more far reaching.

References

Farber, M. J., and P. I. Peterkin. 2000. *Listeria monocytogenes*. *In* The microbiological safety and quality of food, (Eds) Lund, B.M., Baird-Parker, T.C. and Gould, G.W. Aspen Publishers, Gaithersburg, MA., 1178-1232.

Lashkari, D. A., J. L. DeRisi, J. H. MCCusker, A. F. Namath, C. Gentiles, S. Y. Hwang, P. O. Brown, and R. W. Davis. 1997. Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc. Natl. Acad. Sci. USA 94 : 13057-13062.

McGonigle, B., S. J. Keeler, S. M. Lau, M. K. Koeppe, and D. P. O'Keefe. 2000. A genomics approach to the comprehensive analysis of the glutathione S-transferase gene family in soybean and maize. Plant Physiol. 124 : 1105-20.

Schenk, P. M., K. Kazan, I. Wilson, J. P. Anderson, T. Richmond, S. C. Somerville, and J. M. Manners. 2000. Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. Proc. Natl. Acad. Sci. USA 97 : 11655-60.

Yazaki, J., N. Kishimoto, K. Nakamura, F. Fujii, K. Shimbo, Y. Otsuka, J. Wu, K. Yamamoto, K. Sakata, T. Sasaki, S. Kikuchi. 2000. Embarking on rice functional genomics via cDNA microarray: use of 3' UTR probes for specific gene expression analysis. DNA Res. 7 : 367-70.