



Microarray Technologies – An Overview

The University of California San Diego Extension, Bioscience, Microarray Technologies – An Overview, March 13–15, 2002

Gary Hardiman

Biomedical Genomics Microarray Facility (BIOGEM), Division of Biology, University of California San Diego, 9500 Gilman Drive, La Jolla CA 92093-0349, USA

Tel: +1 858 822 3792; Fax: +1 858 822 3021; E-mail: ghardiman@ucsd.edu

The confluence of biotechnology, computer sciences and the completion of genome sequencing efforts for several organisms have led to revolutionary changes in biomedical research. Genomics and informatics are playing increasingly important roles as discovery tools in the basic biological sciences and as diagnostic and rational therapeutic aids in the clinical arena. In recent years, the synthesis or fabrication of high-density arrays of specified DNA sequences comprising all known genes of an organism on a single glass slide or 'chip' have gained popularity. Labeled RNA or DNA targets (such as messenger RNAs (mRNAs) obtained from cells, tissues or organisms under different conditions) can be analyzed by hybridization on DNA [1-4]. This technology allows several types of questions to be asked on a qualitatively different scale than has been previously possible and will have a dramatic impact on the pace of discovery in the field of molecular medicine and drug discovery.

The drug discovery process seeks to develop a biological or chemical entity that, when administered to a patient, will improve the disease symptomatology or actually treat the underlying pathophysiological basis of the particular disease state. A target is the biological entity, usually a gene, mRNA or protein, with which a pharmaceutical is designed to interact. The completion of the human genome sequence has defined the genomic relationships between drug targets. In addition,

sequencing of pathogenic lower organisms has uncovered a host of novel microbial targets.

Gene expression profiling has been a mainstay molecular biological approach for many years. The availability of complete genome sequence information along with improved technology has created a situation where these techniques can be applied in a much broader manner, thereby facilitating profiling of large sections of the transcriptome. The University of California San Diego Extension, Bioscience Microarray Technologies – An Overview meeting, March 13–15, 2002 provided an important forum for reviewing advances and future trends in this field (see Box 1).

Conventional array and chip methodologies

DNA microarrays are comprised of a library of genes, immobilized in a grid on a glass microscope slide. Each unique spot or feature on the grid contains a DNA sequence derived from a specific gene that will bind to the mRNA produced by the gene in question. The standard microarray experimental platform consists of comparing mRNA abundance in two different samples. One fluorescent target is prepared from control mRNA and the second from mRNA isolated from treated cells or tissue under investigation. Both targets are mixed and hybridized together on the same microarray slide and target gene sequences hybridize to their comple-

mentary sequences. The microarray is excited, using a laser, to enable the fluorescent intensity of each spot to be determined. The relative intensities of the two colored signals on individual spots are proportional to the amount of specific mRNA transcripts in each sample, enabling an estimation of the relative expression levels of the genes in sample and control populations. The enormous power of DNA microarray technology is also the source of many of its problems. Errors in handling cDNA clones and cross contamination have been well documented [5]. Furthermore inherent systemic variations are caused by the nature of slide chemistry, target labeling, printing and fluctuations in performance of the scanning instrumentation.

In amplifying large cDNA clone collections for the production of microarrays, great care needs to be taken to minimize sample cross contamination and to ensure the viability of each clone. R Rouse (University of California, San Diego, CA, USA) and D Gurevitch (Laboratory Robotics Interest Group, San Diego, CA, USA) discussed the use of automation in microarray laboratories. The use of automation minimizes potential sources of error due to manual intervention. Furthermore, when amplifying thousands of clones, the main objective is to minimize the total number of reactions needed to generate sufficient material for printing. A high yield of pure amplified polymerase chain reaction (PCR) material means that plentiful amounts of DNA are available for spotting, reduction in cost and labor and ultimately the production of greater numbers of arrays per PCR reaction performed. Automation facilitates the achievement of this goal. The application of automation to the microarray experiment reduces hands-on slide processing time and the number of slide handling

Box 1. Recent advances and future trends in microarray technologies.

- Progress has been made in optimizing the reproducibility of DNA microarrays through the use of automation in fabrication and experimentation.
- Affymetrix (Santa Clara, CA, USA) has recently released a human genome GeneChip pair providing genome coverage. Utilization of eleven probe pairs per gene does not result in a loss of accuracy and should result in a decrease in per sample cost.
- Recent progress in combining the use of ChIP assays with DNA microarrays has allowed genome-wide analysis of transcription factor localization to specific promoter sequences in living cells.
- The Nanochip from Nanogen can be used for genotyping and gene expression profiling with DEP. The future electronic chips will be inexpensive diagnostic 'Lab on a Chip' devices.
- RLS technology with its ease of use and high sensitivity will provide a powerful alternative to fluorescence detection methods.
- Protein microarray technology is considerably more complex than DNA based methods and there are still many technical challenges to overcome.

steps required. Furthermore, the volume of solutions used is greatly reduced. The use of an automated slide processor (ASP) has the following features:

- improves experimental reproducibility
- avoids user inconsistencies
- maximizes hybridization efficiencies
- enhances signal intensities

The complementary technology pioneered by Affymetrix (Santa Clara, CA, USA) consists of short single stranded DNA segments, oligonucleotides or oligos, built to order by chemical synthesis. The GeneChip has traditionally packed up to 400,000 different oligos on a single array, usually representing 10,000 genes with 40 oligo features per

gene. MJ Lelivelt (Affymetrix, Santa Clara, CA, USA) provided an overview of the performance of the newly released human U133 array set, which contains probes respectively for 33,000 genes and 39,000 transcripts. This two chip pair, termed U133subA and U133subB, respectively, contains 45,000 probe sets, with each probe set containing 11 probe pairs and each being 18 microns in size. Approximately 6500 probe sets are included on the chips to detect alternative transcripts and ~ 4500 to detect alternative polyadenylation site variants. The new arrays additionally contain a new set of 100 normalization controls, which offer users additional options for normalization and scaling of GeneChip data. Utilization of 11 probe pairs per set does not result in a loss of accuracy and will ultimately represent a decrease in sample cost. All input sequences for these new chips were mapped to the April 2001 genome draft assembly, and only sequences with high quality alignments were used for annotation.

Conventional microarray analyses, fabricated in house, complement the use of commercially available Affymetrix GeneChips. For example, microarrays containing a few hundred to a few thousand cDNA targets of particular interest can be printed at a relatively low cost, allowing multiple experimental conditions to be examined that would be prohibitively expensive using commercially available arrays. Recent progress in combining the use of chromatin immunoprecipitation (ChIP) assays with DNA microarrays has allowed genome-wide analysis of transcription factor localization to specific promoter sequences in living cells [6,7].

Promoter microarrays

Understanding how DNA binding proteins control global gene expression requires a knowledge of the chromosomal locations at which these proteins function. Currently, a methodology termed ChIP is used to determine whether a particular DNA binding protein is bound to the promoter or enhancer of a specific gene in living cells. In this assay, cells are briefly

treated with formaldehyde to crosslink transcription factors to DNA. DNA is then isolated from the cells and sheared to an average size of 300–400 base pairs. DNA fragments with crosslinked proteins are then subjected to immunoprecipitation with either a control antibody or an antibody that is specific for a transcription factor of interest. The immunoprecipitated protein-DNA adducts are then heated to reverse formaldehyde crosslinks and PCR amplification is performed using specific primers to determine whether a particular promoter or enhancer was coprecipitated with the transcription factor. If a specific DNA sequence is enriched in the population of sequences precipitated by the specific antibody, the transcription factor is considered to have been bound to that sequence in the cell. By carrying out crosslinking following hormonal stimulation or under different physiological conditions, it has been possible to characterize dynamic changes in the association of different transcription factors and coregulators with specific genes. The widespread implementation of this assay in the last few years has had a significant impact on the understanding of how specific transcription factors ultimately activate or repress target genes.

In 2000, Ren *et al.*, demonstrated that it was possible to modify this assay to allow genome-wide location analysis using yeast as an experimental organism [6]. To accomplish this, the pool of specifically precipitated DNA was amplified by ligation mediated PCR and labeled with Cy5 dye. In parallel, unenriched DNA fragments were amplified in an identical manner and labeled with Cy3 dye. The two pools of labeled DNA were then mixed and hybridized to a microarray containing all of the intergenic regions of the yeast genome. Following high stringency washing, the microarray was scanned and the ratios of Cy3–Cy5 intensity were determined for each intergenic region on the array. Using an error model especially developed for this purpose, DNA fragments that were significantly enriched by immunoprecipitation were identified. This process identified all of the known genes for the yeast tran-

scription factor Gal4, as well as several new genes that were confirmed by conventional CHIP assays and additional functional assays [6].

Electronic chip technology

An interesting development in recent years has been electronic chip technology. E Weidenhammer from Nanogen (San Diego, CA, USA) reviewed the development and application of electronic microarrays. The NanoChip exploits the charged nature of biological molecules. Electronic charges are used to rapidly move molecules and concentrate them over defined sites on an array. The concentration of biological materials with electronics enables rapid hybridization reactions; instead of the 12–16 h traditionally required for passive hybridization, the electronic hybridization reactions are performed in 2 min. When a test site on the NanoChip is charged electronically, DNA or RNA rapidly move to that site. Other sites, which are not charged, do not attract DNA or RNA. DNA can therefore be moved and concentrated to any given site on the NanoChip by simply charging the appropriate test site. Each site can be individually charged electronically via platinum wires and can contain an individual assay or experiment. Electronic hybridization and stringency is carried out with single base resolution. Several applications exist to date for electronic microarrays, including:

- polymorphism detection
- gene expression profiling
- short tandem repeat (STR) quantitation
- on-chip target amplification and detection

M Heller (University of California and Nanogen, San Diego, CA, USA) and M Madou (Nanogen) predicted that the next generation integrated devices and systems will have utility in the healthcare industry in point of care diagnostics applications and as biosensors for bioterrorism response, and will represent in the true sense a 'Lab on a Chip'. For the US governments fiscal year 2003, which commences in October 2002, President

GW Bush has requested a budget of US\$27.3 billion for the National Institute of Health. About US\$1.5 billion is focused on bioterrorism related research and infrastructure [8]. It is thought that future chips will be allow electronic sample processing and cell separation. These miniature sensors will be inexpensive, thus permitting rapid and reliable genotyping of pathogens and allowing the discrimination of a one base pair difference between microbes. These sensors will need to be accurate, intolerant of false positives and able to detect very low levels of pathogens and endure wide sample variation from air, food and water to contaminated surfaces such as mail. The NanoChip can currently be used for genotyping and gene expression profiling with dielectrophoretic cell separation (DEP). A 2 min preparation time is required for electronic sample separation of *Listeria* from whole blood by dielectrophoresis. This is the prototype sample to answer systems for multiple pathogen detection. The merging of DNA arrays with microfluidics will enable the transition from research instrument to disposable diagnostics. Novel and less expensive detection techniques will be another important contributor to success in molecular diagnostics.

Resonance light scattering technology

A proprietary labeling technology resonance light scattering (RLS) developed by Genicon Sciences (San Diego, CA, USA) has application in DNA sequencing and DNA and protein microarrays. J Yquerabide (Genicon Sciences) discussed the simplicity and ease of use of this method. Low analyte concentrations that usually require expensive and sophisticated instruments for detection can be detected with RLS particles by eye or using a simple filament lamp. The RLS method can be illustrated by illuminating a 60 nM gold particle suspension with a fine beam of white light. The scattered light has a clear, not cloudy, green color and can be detected at particle concentrations as low as 10^{-15} M [9-11].

RLS particles have many advantages as labels for analyte detection. The light

scattering power of a 60 nM gold particle is equivalent to about that of 500,000 fluorescein molecules. In suspension, 60 nM gold particles are detected by the naked eye. Individual gold particles can readily be seen under a basic microscope with simple dark field illumination. RLS particles have single particle detection sensitivity and, unlike cyanine dyes, do not photobleach. The sensitivity of RLS is much higher than that of fluorescence and is comparable or better than radioactivity and the detection time is much faster. The color of the scattered light can be altered, by changing RLS particle composition or size. The particles can be coated with antibodies or DNA probes for detection of specific analyte antigens or DNA sequences.

Ultrasensitive quantitative assays can be conducted with relatively simple instrumentation. On solid surfaces, a single RLS particle is visible to the eye. For use with gene expression microarrays, the RNA target is reverse transcribed and a hapten is incorporated into the cDNA in place of the standard fluorescent dye. cDNA is denatured and treated with RNase, prior to hybridization to a microarray. The array is washed after hybridization and antihapten antibodies coupled to RLS particles are added bound, and detected. RLS particles can additionally be used with Affymetrix GeneChips and have been hybridized to the CYP450 high-density genotyping arrays.

Proteomics and microarrays.

In the strictest sense of the definition of 'proteome', only one proteome exists in each individual, namely the full complement of all proteins encoded by the genome. Now a more common usage of the term 'proteome' appears in the literature and this definition refers to all of the proteins expressed by a particular cell or tissue at a specific time. The quantitative comparison of these proteomes is a major goal of proteomics, analogous to gene expression profiling. Proteomics is therefore a global study of gene expression at the protein level, encompassing comparative, functional and structural proteomics. Comparative proteomics

seeks to quantify the relative abundance of each protein species present in two or more proteomes. Considering that proteins and not mRNA transcripts are the major effector molecules in the cell, the question often posed is why comparative proteomics is not the method of choice for the global analysis of gene expression

B Cravatt (The Scripps Research Institute, La Jolla, CA, USA) addressed this question and discussed the challenges posed by high-throughput manipulation of proteins. There is no methodological equivalent to the Gene-Chip for comparative proteomics at this time. Transcript analysis offers many technical advantages over protein analysis in that the mRNA molecules possess high affinity and specificity binding partners. Additionally, mRNA molecules exhibit equivalent biochemical properties and can be amplified. In contrast to mRNAs there is no PCR equivalent for proteins. Furthermore, proteins do not possess straightforward binding partners and exhibit painfully diverse biochemical features. In spite of these challenges, microarrays are being developed and employed for comparative proteomic studies.

Protein microarrays aim to dramatically increase the amount of proteomic information obtained, as a function of both time and sample quantity, by permitting the parallel analysis of numerous proteins in a miniaturized assay format. Protein affinity ligands are arrayed onto glass slides as capture probes for detection of proteins in complex samples. Several technical hurdles face protein array technology namely the acquisition, arraying, and stable attachment of ligands, to chips and detection of the respective binding partners.

Ligands can be attached to slides by covalent bonding (aldehyde-amine reactions), random orientation or via affinity tags (His6 tag on nickel coated slides). Different chip formats exist, with glass slides, matrix slides and nanowells being popular formats. Glass slides are compatible with standard microarray and detection equipment and are relatively inexpensive, but are prone to high evaporation rates and cross contamination.

Matrix slides and nanowells reduce evaporation and minimize cross contamination. One disadvantage with matrix slides is cost, as they are fabricated by expensive photolithography.

Historically, antibodies have represented the most general type of high affinity, high selectivity protein binding reagent as they can be generated against almost any pure protein. Recombinant antibodies can be produced in both high quantity and purity for arraying purposes. The biggest challenges that faces antibody arrays is obtaining high affinity and high specificity antibodies against the 100,000 or more proteins in the human proteome. Currently, antibodies exist against a mere 2–3% of the proteome. Furthermore the specificity of many of these antibodies is poorly documented and additional antibodies may be required to detect post-translational modifications. A second fluorescently tagged antibody is sometimes used for detection, raising the number of antibodies per arrayed protein to two.

Detection of proteins bound to antibody arrays is more complex than DNA microarrays. Proteomes under comparison can be labeled in an analogous fashion with different fluorophores, but the reproducibility of these chemical reactions and interference with protein-antibody interactions pose additional problems. Other detection methods employed include atomic force microscopy and surface plasma resonance. One disadvantage with antibody arrays is that features of the bound protein, namely size and post-translational modification, such as glycosylation, phosphorylation and acetylation are not seen directly.

Other protein binding molecules, such as aptamers, are being proposed as alternatives or complements to antibodies. Aptamers are protein binding RNA molecules, which are easy to select for, synthesize and array. Purified protein is a requirement for use with aptamers and there is potential for biased binding as RNAs tend to be highly negatively charged. Chemical binding surfaces, such as the protein chip from Cipher-

gen Biosystems (Fremont, CA, USA) are useful alternatives to antibody arrays. These surfaces are nonspecific, facilitating the binding of many diverse proteins. Although the mass spectrophotometric readout is rapid, limited spatial separation of bound proteins can lead to a heavy bias towards the detection of abundant proteins. A recent study successfully applied this technology to analyze proteomic patterns in serum and facilitate the correct identification of ovarian cancer samples. This technology can be used as a screening tool for all stages of ovarian cancer in high-risk and general populations [12].

Protein microarrays have been employed for the global analysis of the yeast proteome. Zhu *et al.* expressed and arrayed 5800 different yeast proteins as His6 tagged fusion proteins on nickel-coated slides. Minimal quantities of protein were required (each spot contained 10–1000 fg of protein). Utilizing this array format, researchers can screen for an activity that provides a visual readout, such as radioactivity, fluorescence or chemiluminescence, thereby identifying new protein functions in a relatively unbiased manner. The disadvantage with this format is that the arrays are relatively biased in terms of expression. Few membrane or secreted proteins will be functionally expressed using this format.

One issue posed by B Cravatt is that there is not always a direct correlation between protein abundance and activity. Many proteins, for example proteases, kinases and phosphatases are synthesized and secreted as inactive zymogens and are converted to an active form by enzymatic cleavage. Additionally, one has to remember with protein microarray analysis that proteins are expressed and analyzed in artificial environments relative to the cell and these environments may not have the same physiological relevance. Nevertheless, it can be expected that in the coming years strides and advances in protein array technology will be seen but challenges to this field are predominantly fundamental problems posed by protein biochemistry and not issues of chip surface, chemistry or engineering.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Brown PO, Botstein D: Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21, 33-37 (1999).
2. Cheung VG, Morley M, Aguilar F *et al.*: Making and reading microarrays. *Nat. Genet.* 21, 15-19 (1999).
3. Debouck C, Goodfellow PN: DNA microarrays in drug discovery and development. *Nat. Genet.* 21, 48-50 (1999).
4. Duggan DJ, Bittner M, Chen Y *et al.*: Expression profiling using cDNA microarrays. *Nat. Genet.* 21, 10-14 (1999).
5. Knight J: When the chips are down. *Nature* 410, 6831 (2001).
- **The enormous power of DNA microarray technology is also the source of many of its problems. Errors in handling cDNA clones and cross contamination problems are discussed in this article.**
6. Ren B, Robert F, Wyrick J *et al.*: Genome-wide location and function of DNA-associated proteins. *Science* 290, 2306-2309 (2000).
- **Description of the genome-wide location process, which identified all of the known genes for the yeast transcription factor Gal4, as well as several new genes that were confirmed by conventional ChIP assays and additional functional assays.**
7. Ren B, Cam H, Takahashi Y *et al.*: E2F integrates cell cycle progression with DNA repair, replication and G(2)/M checkpoints. *Genes Dev.* 16, 245-256 (2002).
8. Agres T: Bioterrorism projects boost US research budgets. *The Scientist* 16(6), 26-27 (2002).
9. Yguerabide J: Theory for establishing proximity relations in biological membranes by excitation energy transfer measurements. *Biophys J.* 66(3 Part 1) 683-693 (1994).
10. Yguerabide J, Yguerabide EE: Light-scattering submicroscopic particles as highly fluorescent analogs and their use as tracer labels in clinical and biological applications: experimental characterization: theory. *Anal. Biochem.* 262, 137-156 (1998).
11. Yguerabide J, Yguerabide EE: Light-scattering submicroscopic particles as highly fluorescent analogs and their use as tracer labels in clinical and biological applications: Experimental characterization: Experimental characterization. *Anal. Biochem.* 262, 157-176 (1998).
- **RLS particles have many advantages as labels for analyte detection.**
12. Petricoin EF 3rd, Ardekani AM, Hitt BA *et al.*: Use of proteomic patterns in serum to identify ovarian cancer. *The Lancet* 359, 572-577 (2002).
- **Application of Ciphergen protein chips to analyze proteomic patterns in serum and facilitate the identification of ovarian cancer samples. This technology can be used as a screening tool for all stages of ovarian cancer in high risk and general populations.**
13. Zhu H, Bilgin M, Bangham R *et al.*: Global analysis of protein activities using proteome chips. *Science* 293, 2101-2105 (2001).
- **An elegant study where protein microarrays were utilized for global analysis of the yeast proteome.**