Research review paper

cDNA microarray technology and its applications

Charlie C. Xiang\textsuperscript{a,*,1}, Yidong Chen\textsuperscript{b,2}

\textsuperscript{a}Department of Molecular Virology, Virology Division, US Army Medical Research Institute of Infectious Diseases, Ft. Detrick, Frederick, MD 21702, USA

\textsuperscript{b}National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

The cDNA microarray is the most powerful tool for studying gene expression in many different organisms. It has been successfully applied to the simultaneous expression of many thousands of genes and to large-scale gene discovery, as well as polymorphism screening and mapping of genomic DNA clones. It is a high throughput, highly parallel RNA expression assay technique that permits quantitative analysis of RNAs transcribed from both known and unknown genes. This technique provides diagnostic fingerprints by comparing gene expression patterns in normal and pathological cells, and because it can simultaneously track expression levels of many genes, it provides a source of operational context for inference and predication about complex cell control systems. This review describes this recently developed cDNA microarray technology and its application to gene discovery and expression, and to diagnostics for certain diseases. © 2000 Published by Elsevier Science Inc. All rights reserved.

Keywords: cDNA microarray; Technology; RNA expression assay technique

1. Introduction

To transform a single fertilized egg cell into an adult human body and then keep that body alive and healthy requires about 100 000 genes, each of which must adjust its expression to precise degrees and at precise times and locations. Three billion base pairs of the human genome will be completely sequenced by the year 2003. The next great challenge is to find out how those genes act in concert to regulate the whole organism. To understand the development of organisms, the onset of genetic diseases, or the concerted functions of genes in regul-
lating cell function or transformation, the best procedure is to monitor the fluctuating activi-
ties of the genes in different tissues at different stages of development, in good and in bad
health. However, the complete mechanism or correlation among the genes has not yet been
elicted for even a single cellular function. Monitoring the activities of a panel of genes to de-
terminate the role of genes in regulating any biological process in a whole organism is a formi-
dable job. Until 3 years ago it was only possible to tackle one gene at a time. The DNA mi-
croarray technique offers the best way to approach such a daunting task. This technology has
been steadily developed for more than 3 years since Patrick Brown and his colleagues first
published their work in October 1995 [1]. It seems likely to become a standard tool of both
molecular biology research and clinical diagnostics.

There are basically two types of microarrays that have been developed so far: cDNA mi-
croarray and oligonucleotide array. Arraying methods include on-chip photolithographic
synthesis of 20–25 mer oligos onto silicon wafers [2], off-set printing of 20–25 mer oligos
array, and 500–5000 bp cDNAs printed onto either glass slides or membranes [3]. An ideal
support allows effective immobilization of a target onto its surface, and robust hybridization
of probe with the target. Glass has many of the same advantages as nylon, which is the other
standard support used for making microarrays. It also has unique advantages including cova-
lent attachment of DNA samples onto a treated glass surface; durable material that sustains
high temperatures and washes of high ionic strength; nonporous support so the hybridization
volume can be kept to a minimum, thus enhancing the kinetics of annealing probes to targets;
no significant contribution to background noise because of its low fluorescence; and hybrid-
ization of the array with two or more probes labeled with different fluors for serial or parallel
analyses [4]. This review will focus on cDNA microarrays on glass and its applications.

2. Biological aspects of cDNA microarray technology

Life depends on the ability of cells to store, retrieve, and translate the genetic instructions
required to make and maintain a living organism. Hereditary information is passed on from a
cell to its daughter cells at cell division, and from generation to generation of organisms
through the reproductive cells. These instructions are stored within every living cell as the
genes—the information-containing elements that determine the characteristics of a species as
a whole and of the individuals within it. A cell relies on its gene products for a wide variety
of functions including energy production, macromolecule biosynthesis, cellular architecture
maintenance, and response to environmental stimuli. Proteins are the active working compo-
nents of the cellular machinery, whereas DNA stores the information for protein synthesis
and RNA carries out the instructions encoded in DNA. Expression of the information in
DNA is mediated by RNA. Messenger RNAs (mRNAs) are the transcripts that carry the spe-
cific information for the sequence of amino acids in proteins. There are very many mRNA
species of widely varying sizes, generally possessing little secondary structure. With rare ex-
ception, all species of eukaryotic mRNAs are polyadenylated. The amount of mRNA in
mammalian cells has been estimated at approximately 500 000 mRNA molecules per cell.
They are relatively stable in eukaryotes. The half-life of mRNAs in mammalian cells can last
for days. Generally speaking, the level of a mRNA in a particular cell represents the meta-
bolic activity of the specific gene. The human cell contains about 50,000–100,000 genes. Some genes are expressed in all cell types. Others are only expressed in particular cell types. As cells change their status during their development or the cells are stimulated by physical and chemical reagents, the gene expression patterns in the cells will be altered. For example, the optimal intracellular environment for viral replication is different from the optimal environment for normal cell growth and replication during viral infection. The virus itself has evolved the means to alter the intracellular environment to suit its own ends. Some of these alterations in the intracellular environment result from the changes in the concentrations of cellular RNAs. Comparison of gene expression patterns by measurement of mRNA levels in normal and pathological cells could provide useful diagnostic ‘fingerprints’ and help identify aberrant functions that would be reasonable targets for therapeutic intervention.

3. Bioinformatics in cDNA microarray technology

During the early stages of microarray technology, much of the effort went into the development of the enabling technology, such as serial analysis of gene expression (SAGE) [5,6], oligonucleotide arrays, and cDNA arrays. At the same time, relatively little attention had been paid to data analysis methods and interpretation of gene expression. One of the reasons is due to data insufficiency of the gene expression pattern. The same situation was endured in the early stage of DNA sequence database development. Only after more than 15 years of accumulation of DNA sequences in GenBank were many of sequence-based tools like BLAST developed. With the introduction of microarray technology, geneticists face the rapid evolution of bioinformatics driven not only by the problems associated with gene mapping and sequencing, but also by the massive quantities of data generated by parallel expression monitoring technology. Currently, with the advances of the genome-scale technologies, the biomedical research publications have shifted their attention to the functional aspects of thousands of genes, instead of a single gene’s structural information. However, the bioinformatics community still faces many of the challenges posed by the parallel approaches of gene expression study. Some of most pressured subjects include: (1) laboratory information management, (2) microarray image analysis, (3) gene expression profiling analysis, (4) genetic fingerprint or expression pattern classification, and (5) other high-level control system modeling issues.

3.1. Laboratory information management system

The main goal of laboratory information management system (LIMS) is to track and manage material information flow [7]. The ArrayDB (http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/) was developed to store, retrieve, and analyze microarray experiment information. For each microarray experiment, the basic information in the database consists of: (1) information about microarrays and specific cDNA clone inserts (such as clone ID, title, database hyperlink, etc.); (2) information about probes (such as sample source, experiment condition, fluoro-type, etc.); and (3) information about images (raw scanned image, intensities and ratios for each target clone, etc.). The other goal of ArrayDB is to identify patterns and relationships among intensity ratios, both individual and across multiple experiments. In addition,
the option to download data associated to microarray experiment over the Internet adds flexibility to the end-users’ customized data analysis.

3.2. Microarray image analysis

The objective of microarray image analysis is to extract probe intensities or ratios at each cDNA target location, and then cross-link printed clone information so that biologists can easily interpret the outcomes and perform further high-level analysis. However, the microarray image sources are not only from one print-mode (i.e. different printing tip arrangement or different arrayers) [8] or one hybridization method (i.e. fluorescent [Stanford, NIH, etc], radioactive probe, and others) [9], the analysis methods are very different. Typically, microarray image analysis consists of cDNA target segmentation, target detection, local background intensity and probe fluorescent intensity measurement, ratio analysis, and, finally, data visualization. The keys to success of a single slide analysis are the measurement of expression ratio between samples and the selection of a set of ‘housekeeping’ genes for ratio calibration.

3.3. Gene expression profiles and fingerprinting

Single microarray hybridization processed by microarray image analysis software may be sufficient for much high-throughput gene expression screening. For some other experiments, such as the study of development of some biological systems, monitoring the change of temporal gene expression patterns becomes very important to study the biological interactions between genes within an organism [10], to understand the mechanism of some orderly behavior of gene expressions [11], and to look for relatedness among cell types [12].

The first step of data analysis is often referred to as ‘data exploration,’ in which any nonrandom patterns or structures requiring further explanation are recognized. Clustering is one of the techniques in which the data of interest are placed into a small number of homogeneous groups or clusters. To study the orderliness of expression data, the multidimensional scaling (MDS) technique is routinely used for biological sample similarity visualization. For every gene used in the microarray experiments, K-mean-based algorithms or hierarchical clustering methods (similar to Eisen et al. [13]) are used for gene expression similarity clustering and visualization.

4. Making cDNA microarray

Fig. 1 describes the basic principles of the DNA microarray technology. Microarrays containing large DNA fragments, such as cDNAs, are generated by physically depositing small amounts of each DNA of interest onto known locations on glass surfaces. They begin with the selection of the targets to be printed on the array. Deposition strategies typically produce microarrays consisting of groups of cDNAs amplified by the polymerase chain reaction (PCR) ranging from 0.5 to 2.0 kb. In many cases, the targets are chosen directly from databases such as GenBank, dbEST, and UniGene. Additionally, full-length cDNAs, collections of partially sequenced cDNAs (or ESTs), or randomly chosen cDNAs from any library of interest can be used. The DNA fragments were PCR amplified from individual clone using specific primers or universal primers if all the genes were cloned in the universal vector.
The microarrays were fabricated on poly-L-lysine (Sigma, St. Louis, MO, USA) coated microscope slides. The DNA fragments are cross-linked by UV to the matrix. After fixation, residual amines on the slide surface react with succinic anhydride to reduce the positive charge at the surface.

A robotic arraying machine loaded about 1 μL of PCR-amplified fragments from corresponding wells of 96-well plates and deposited about 5 nL of each sample onto each of 42–110 slides. Several commercial arraying machines are currently used including MicroGrid from BioRobotics (Cambridge, UK), GMS 417 from Genetic Microsystems (Woburn, MA, USA), OmniGrid from GeneMachines (San Carlos, CA, USA), and the PixSys PA series from Cartesian Technologies (Irvine, CA, USA). The arrayer made by Beecher Instruments (Silver Spring, MD, USA) in the National Human Genome Research Institute at National Institute of Health in Bethesda, Maryland comprises an xyz cantilever type robot holding 16 quill pen-type probes, a removable vacuum chuck for 48 standard microscope slides, a 20 microtiter tray loader/stacker, a wash/dry station, a controlling PC, air-handling components, and a cabinet (Fig. 2). The robot moves only the probe holder. Sixteen stainless steel probes, which function as ‘quill pens,’ are spaced on 9-mm centerlines to conform to the well spacing of standard 96-
well microtiter trays. The probes are spring loaded to accommodate small differences in probe length and slide and well positions. The PC controls all of the other components and allows operator input of various parameters, such as the number of probes, trays, and slides, spot spacing, and pattern on the slides, duration of each cycle component and speeds and accelerations of the robot. The arrayer is able to put down 16 spots on each of 48 slides and to wash and dry the probes for the next set of cDNAs for the next tray in about 70 s. Most of this time is taken up with the actual spotting, as the wash and dry cycles are about 2 s each and the loading is about 10 s. Thus the contents of one 96-well tray can be spotted every 7 min, and 10 000 spots should take about 12 h. Many arraying machines use the ‘quill’-type spotting tips that were originally designed by Pat Brown and his colleagues at Stanford University. But other companies developed solid pins and the GMS 417 from Genetic Microsystems uses a special device, so-called Pin-Ring (Fig. 3).

5. Probe labeling and array hybridization

Total RNA is typically isolated from different tissues and cell lines using the Trizol reagent from BRL according to the manufacturer’s instructions. The purity of RNA is a crit-
ical factor in hybridization performance, particularly when using fluorescence, since cellular protein, lipid, and carbohydrate can mediate significant nonspecific binding of fluorescently labeled cDNAs to slide surfaces. Several total RNA extraction kits have been tested to produce satisfactory results, including Trizol reagents from BRL and RNeasy kits from Qiagen.

Fluorescently tagged cDNA probes from the pair of samples to be analyzed were pooled together. The block reagents, such as yeast tRNA, poly dA, and human Cot-1 DNA, were added. The final hybridization solution was $3 \times SSC$ and 0.1% sodium dodecyl sulfate (SDS). The probes were hybridized to an array at 65°C for 16–24 h. Unbound probes were removed by washing in SSC and SDS solutions at room temperature. The slides were centrifuged to remove residual liquid and were air-dried. For the detailed protocols for probe labeling and array hybridization, please refer to the web site of Cancer Genetics Branch in the National Human Genomic Research Institute at http://www.nhgri.nih.gov/DIR/LCG/15K/HTML.

6. Array scanning and data analysis

To determine which DNAs correlate with changes in gene expression or toxic effects, the microarrays are first scanned to produce visual images and to generate raw numerical data for each spot on the array. The microarray reader is basically a computer-controlled inverted scanning fluorescent confocal microscope with a double or multiple laser illumination system, such as ScanArrayer 4000 and 5000 from General Scanning (Watertown, MA, USA), Avalanche from Molecular Dynamics (Sunnyvale, CA, USA), GMS 418 from Genetic MicroSystems (Woburn, MA, USA), and GeneTAC from Genomic Solutions (Ann Arbor, MI, USA). In addition, some companies, such as Genometrix (The Woodlands, TX, USA; Applied Precision, Seattle, WA, USA), are developing charge-coupled device (CCD) cameras to capture the microarray images. The illumination system of the NHGRI Scanner manufactured by Beecher Instruments (Silver Spring, MD, USA) consists of three air-cooled lasers.

Fig. 3. Pin-and-ring array technology by Genetic MicroSystems. The pin-and-ring array technology is capable of creating spots of extremely consistent size, shape, and volume (Courtesy of Genetic MicroSystem.)
(Fig. 4): a 488-nm, 100-mw Argon ion laser for exciting FITC; a 532-nm, 100-mw NdYag for Cy3; and a 633-nm, 35-mw HeNe for Cy5. Any two lasers may be turned on simultaneously and their beams are delivered to the specimen via a single dichroic and an objective lens (0.75 NA, 0.66-mm width). The emitted light, after passing back through the objective and primary dichroic, is focused through a confocal pinhole and through a secondary dichroic onto two cooled photo multiplier tubes (PMTs), which operate in parallel for the two different wavelengths. The stage is a standard computer-controlled microscope stage capable of 100 mm/s scans and 5 micron resolution. One or two standard 25 × 75 mm slides can be scanned at a time. At 100 mm/s, with 20 micron pixels, an area of 40 × 20 mm (capable of printing easily about 10 000 spots) can be scanned in about 20 min. About 10 pg/μL of each species of cDNA can be reliably detected. Fig. 5 shows the images from two probes after scanning and a combined color image.

Microarray image analysis programs, ArraySuite, as a set of extensions for IPLab Spectrum for Macintosh computer (Scanalytics, Inc., Fairfax, VA, USA) and developed at NHGRI, are used to analyze red and green hybridization intensities and red/green ratio for every gene on the array [9]. Features of ArraySuite include array alignment, DeArray, and Target-Locator. Fig. 6 illustrates the entire process of image analysis.
7. DNA microarray applications

7.1. Use of DNA microarrays for gene expression and discovery

Measuring transcript levels for thousands of genes in parallel is one of the more widespread applications of DNA microarray technology. Schena et al. [1] first described the high-capacity system of cDNA microarrays to monitor the expression of 45 Arabidopsis genes in parallel. Since then the cDNA microarray applications have been reported in many organisms, including plant [14], yeast [3,15], and human beings. The cDNA microarray technology was used to profile complex diseases and discover novel disease-related genes. DeRisi et al. [16] used the cDNA microarray technique to analyze gene expression patterns in human cancer. They demonstrated the tumorigenic properties of human melanoma cell line UACC-903 could be reversed by insertion of human chromosome 6. Heller et al. [17] studied gene expression characteristic of the inflammatory disease rheumatoid arthritis and inflammatory bowel disease. They documented the stability of the cDNA microarray technology for profiling diseases and for identifying disease-related genes. The technology provided new targets for drug development and disease therapies, and in doing so allowed for improved treatment of chronic diseases that were challenging because of their complexity. Welford et al. [18] used representational difference analysis coupled to cDNA microarray hybridization to de-
tect differentially expressed genes in primary tumor tissue. Their data showed that the use of representational difference analysis essentially provided an enriched library of differentially expressed genes, while analysis of the library with microarray technology allowed rapid and reproducible screening of thousands of DNA molecules simultaneously. They further indicated that the coupling of two techniques in their system resulted in a large pool of differentially expressed genes.

7.2. Use of DNA microarrays for predicting various biochemical pathways

Schena et al. [19] monitored the expression of 1046 human cDNAs of unknown sequence using two-color differential expression analysis of heat shock and phorbol ester-regulated genes. They were the first to demonstrate the ability of cDNA microarrays to rapidly provide data for correlation of gene expression to biochemical pathways. The cDNA microarrays containing virtually all of the genes of *Saccharomyces cerevisiae* have been fabricated. DeRisi et al. [15] studied the metabolic and genetic control of gene expression on a genomic scale in yeast by using the yeast microarrays. They examined the effects of the diauxic shift from anaerobic to aerobic metabolism under glucose limitation and the concomitant switch to ethanol as a carbon source. The significance of their study was that it mapped the changes in expression of genes with known function to their metabolic pathways and vividly showed which metabolic pathways were programmed by the shift. Spellman et al. [11] comprehensively identified cell cycle-regulated genes in yeast by cDNA microarray hybridization. They found 800 genes that meet an objective minimum criterion for cell cycle regulation and half of them respond to either G1 cyclin or B-type cyclin or both of these cyclins. Lyer et al. [20]
studied temporal transcriptional program in the response of human fibroblasts to serum. They applied microarrays with 8600 different human genes to the study of growth control and cell cycle progression in humans. They clustered genes into groups on the basis of temporal patterns of expression of those genes. Their data indicated that many features of the transcriptional program appeared to be related to the physiology of wound repair, which suggested that the fibroblasts play a larger and more important role in the complex multicellular response than had previously been appreciated. Galitski et al. [21] used microarray-based gene expression analysis to identify genes showing ploidy-dependent expression in isogenic Saccharomyces cerevisiae strains that varied in ploidy from haploid to tetraploid. They found that those genes were induced or repressed in proportion to the number of chromosome sets, regardless of the mating type.

7.3. Use of DNA microarrays for drug discovery and development

The expression pattern of a gene provides indirect information about function. Knowledge of highly selective gene expression, as well as sequence homology to a known gene family, could provide a convenient shortcut for implicating a target in a given pathway or disease. The cDNA microarray technology has helped many pharmaceutical companies to identify appropriate targets for therapeutic intervention. DNA microarrays have also been used to monitor changes in gene expression in response to drug treatment. Pietu et al. [22] identified novel gene transcripts preferentially expressed in human muscles by quantitative hybridization of a high density cDNA array. Among those genes, cathepsin K is a novel cysteine protease that is expressed selectively in osteoclasts. The discovery led to the development of drugs to inhibit the cathepsin K. Marton [23] performed drug validation studies and identified secondary drug target effects using DNA microarrays.

DNA microarray or DNA chip technology is a powerful new approach for simultaneously monitoring the relative expression of a large number of genes in a quantitative fashion. cDNA microarray technology is rapidly advancing. Its applications to gene discovery, gene expression, and mapping have been convincingly demonstrated. The suitability of the cDNA microarray for profiling diseases and for identifying disease-related genes has been also well documented. This novel technology could provide new targets for drug development and disease therapies. It may thus provide a useful link between gene sequences and clinical medicine for both human beings and animals. DNA microarrays seem likely to become a standard tool for both molecular biology research and clinical diagnostics.

References