BioFeature

A Concise Guide to cDNA Microarray Analysis

P. Hegde, R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J.E. Hughes, E. Snesrud, N. Lee and J. Quackenbush
The Institute for Genomic Research, Rockville, MD, USA

ABSTRACT

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. We have optimized all aspects of the process, including PCR amplification of target cDNA clones, microarray printing, probe labeling and hybridization, and have developed strategies for data normalization and analysis.

INTRODUCTION

Recently, a variety of techniques including serial analysis of gene expression (SAGE) (12), differential display (5), oligonucleotide arrays (6) and cDNA microarrays (8) have been developed that allow mRNA expression to be assessed on a global scale and the parallel assessment of gene expression for hundreds or thousands of genes in a single experiment. The most common use of these techniques is to determine patterns of differential gene expression or to compare differences in mRNA expression levels between identical cells subjected to different stimuli or between different cellular phenotypes or developmental stages.

Microarray expression analysis (8) has features that make it the most widely used method for profiling mRNA expression. DNA segments representing the collection of genes to be assayed are amplified by PCR and mechanically spotted at high density on glass microscope slides using relatively simple *x-y-z* stage robotic systems to create a microarray containing thousands of elements. Microarrays containing the entire set of genes from a microbial genome or tens of thousands of eukaryotic cDNA clones can be easily constructed. The microarrays are queried in a co-hybridization assay using two or more fluorescently labeled probes prepared from mRNA from the cellular phenotypes of interest (10). The kinetics of hybridization allows relative expression levels to be determined based on the ratio with which each probe hybridizes to an individual array element. Hybridization is assayed using a confocal laser scanner to measure fluorescence intensities, which allows the simultaneous determination of the relative expression levels of all the genes represented in the array.

Efficient expression analysis using microarrays requires the development and successful implementation of a variety of laboratory protocols and strategies for fluorescence intensity normalization. The process of expression analysis can be broadly divided into three stages: *(i)* array fabrication; *(ii)* probe preparation and hybridization; and *(iii)* data collection, normalization and analysis. Here, we present protocols that we have standardized and used regularly in our laboratory for microarray analysis. The procedures described have been tested and refined over the past year and have been optimized using hybridization of RNA derived from cell lines to give reproducible and consistent results. Note that a number of alternative protocols have been published (3) or are available on the World Wide Web (Table 1), but the system described here offers advantages over such protocols. In particular, we have derived a combination of printing, labeling and hybridization conditions that allows a significant reduction in the quantity of starting total RNA required for analysis.

ARRAY FABRICATION

Microarrays are constructed by arraying PCR-amplified cDNA clones or genes at high density on derivatized glass microscope slides. For the analysis of expression in most eukaryotes, expressed sequence tag (EST) data represent the most extensive data for gene identification. ESTs are single-pass, partial sequences of cDNA clones and are used extensively for gene discovery and mapping in humans and other organisms. The EST approach has been widely adopted; more than 71% of all GenBank® entries and 40% of the individual nucleotides in the database are derived EST sequences (9).

Generally, cDNA clones are selected to represent as many unique transcripts as possible. There are numerous analyses of these data that attempt to identify unique human transcripts within the EST data; the two most widely used are UniGene (1) (http://www.ncbi.nlm.nih.gov/UniGene/) at the National Center for Biotechnology Information (NCBI) and The Institute for Genomic Research (TIGR) Human Gene Index (7) (HGI) (http://www.tigr.org/tbd/hgi/hgi.html). While both UniGene and HGI are based on EST clustering, the TIGR protocol assembles the ESTs within the clusters to produce tentative human consensus (THC) sequences.

We selected cDNA clones for array construction using the TIGR HGI as part of a program to assemble a 30 000-gene clone set. THCs were chosen for representation in the clone

BioFeature

National Human Genome Research Institute	http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/protocol.html
Stanford University	http://cmgm.stanford.edu/pbrown/protocols/index.html
Telechem	http://arrayit.com/DNA-Microarray-Protocols/
University of Pennsylvania	http://genomics.med.upenn.edu/vcheung/protocols.htm
The Institute for Genomic Research	http://www.tigr.org/tdb/microarray

set, with preference given to those containing known genes or those with mapped positions; additional THCs were selected to represent as yet uncharacterized transcripts. For each target THC, a single cDNA clone was identified based on the EST content of the THC assembly.

PCR Amplification and Clone Preparation

The cDNA clone inserts can be amplified by PCR from plasmid DNA or directly from clones in culture. In high-throughput applications, amplification of clones from culture has the advantage of being both more cost efficient and less labor intensive with lower cross-contamination rates than amplification from plasmid DNA. Our amplification success rate from culture is equivalent to that achieved using plasmid templates. For more than 30 000 clones, our success rate for single-band amplification is approximately 87.5%; 6.3% of reactions yield multiple or weak bands and 6.2% fail to amplify.
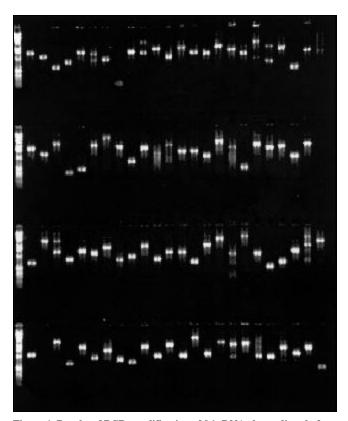

Figure 1. Results of PCR amplification of 96 cDNA clones directly from bacterial culture. The leftmost lane in each tier of the gel contains a 1-kb ladder; each tier contains 24 distinct samples.

Results from a typical amplification using the protocol described here are shown in Figure 1.

PCR Amplification

The cDNA clones that are widely available through the IMAGE consortium distributors—The American Type Culture Collection (ATCC), Research Genetics and Genome Systems—have been cloned into a variety of vectors. While the majority have both M13(-21) and M13REV priming sites, many have point mutations in either of these two universal priming sites. We have designed alternative M13 primers that avoid these point mutations and have allowed amplification of clone inserts from all of the vectors encountered to date. These new universal amplification primers are:

M13 FWD: 5′-GTTTTCCCAGTCACGACGTTG-3′ and
M13 REV: 5′-TGAGCGGATAACAATTTCACACAG-3′.

Clone inserts are amplified using the following protocol:

1. Selected clones are inoculated into 96-well blocks (Qiagen, Valencia, CA, USA) containing 1.2 mL LB/Ampicillin (50 µg/mL) and incubated for 16 h at 37°C and 200 rpm in a shaking incubator. An aliquot of 100 µL is archived for future use in microplates containing 10% glycerol at -80°C.

2. Following overnight growth, 5 µL culture suspension are transferred into a 96-well Falcon® U-bottom plate (BD Biosciences, San Jose, CA, USA) containing 95 µL Milli-Q®-treated water (Millipore, Bedford, MA, USA).

3. Microtiter plates containing diluted culture are heated in a laboratory oven to 95°C for 10 min to lyse the cells and release the plasmid clones.

4. Before PCR, cellular debris is removed by centrifugation at $1200 \times g$ for 3 min in a centrifuge equipped with microplate carriers.

5. Clone inserts are amplified in 50-µL reactions in 96-well reaction plates (PE Biosystems, Foster City, CA, USA). A reaction master mixture is prepared for each reaction plate:

Milli-Q water	3.2 mL
10× PCR buffer	500 µL
$MgCl_2$ (25 mM)	500 µL
M13 Forward primer (10 µM)	100 µL
M13 Reverse primer (10 µM)	100 µL
dNTP mixture (2 mM per dNTP)*	500 µL (Life Technologies, Rockville, MD, USA)
Platinum Taq (5 U/µL)	20 µL (Life Technologies)
Total:	4.92 mL

(*10 µL each dNTP as 100 mM stock in 460 µL water)

6. For each clone, add 48 µL master mixture to 2 µL culture supernatant in a 96-well PCR plate.

7. Reactions are amplified in a thermocycler (MJ Research, Waltham, MA, USA) using the following cycling protocol:

95°C × 2 min		Initial Denaturation
95°C × 30 s		Denaturation
52°C × 30 s	× 30 cycles	Annealing
72°C × 2 min		Extension
4°C forever		

Reaction Clean-Up

For efficient binding of the amplified clone inserts to the slides, it is essential to remove unincorporated nucleotides and primers from the reaction products. While there are various techniques that can be used, we have found filtration using 96-well multiscreen filter plates (Millipore) to give excellent DNA product recovery with no significant contamination at relatively low cost, using the following filtration protocol:

1. PCR product (50 µL) is transferred to the Millipore filter plate.

2. The filter plate is placed on a vacuum manifold filtration system (Qiagen or Millipore) and filtered at a pressure of 15 in (380 mm) Hg for 10 min or until the plate is dry.

3. Water (30 µL) is added to each well and filtered at 15 in (380 mm) Hg for 5–10 min or until the plate is dry.

4. Step 3 is repeated.

5. The plate is removed from the manifold filtration system; 60 µL water is added to each well. The plate is placed on a shaker and shaken vigorously for 10 min to resuspend the DNA.

6. The purified product is manually pipetted to a new 96-well plate.

7. Plates containing the purified PCR products are then sealed using a cap mat (VWR, West Chester, PA, USA) and stored at 4°C for future arraying.

Array Printing

Microarrays are prepared by printing amplified PCR products suspended in either a high-salt or other denaturing buffer to poly-L-lysine- or aminosilane-coated glass microscope slides using a high-speed robotic system. This process, originally described by Patrick Brown and collaborators (8) at Stanford University, is available so that others can replicate their arraying robot (http://cmgm.stanford.edu/pbrown/mguide/index.html)). However, some companies are selling robotic systems for microarraying (Table 2). We use a microarray robot built by Intelligent Automation Systems (IAS) of Cambridge, MA, USA. Based on a high-precision, four-axis Seiko robotic arm, the IAS arrayer uses a 12-tip print head to array DNA samples from either 96- or 384-well microplates to as many as 100 silanized glass microscope slides. With an average spot size of 130 µm and the capability to adjust the spot-to-spot spacing, the IAS arrayer can spot 19 200

Table 2. Manufacturers of Microarray Spotting Robots

Beecher Instruments
http://www.beecherinstruments.com
BioRobotics
http://www.BioRobotics.com/
Cartesian Technologies
http://www.cartesiantech.com/
Engineering Services
http://www.ESIT.com/
Genetic Microsystems
http://www.geneticmicro.com
Genetix
http://www.genetix.co.uk/
Gene Machines
http://www.genemachines.com
Genomic Solutions
http://www.genomicsolutions.com/
Intelligent Automation Systems
http://www.ias.com
Packard
http://www.packardinst.com/

elements (the contents of 200 microplates) or more onto a single slide.

Both the slide surface and the spotting buffer are critical components for reproducible, high-fidelity micorarray analysis. Most published reports have used high-salt buffers (such as 3× SSC) to print DNA on poly-L-lysine-coated slides (10). However, our analysis suggests that aminosilane offers a more consistent surface with lower background fluorescence. There are numerous commercial vendors for aminosilane-coated slides (Table 2), but CMT-GAPS™ (Corning, Corning, NY, USA) aminosilane-coated glass microscope slides have been the most consistent. Using our protocol in a number of side-by-side comparisons, these slides produce approximately half the background fluorescence of poly-L-lysine slides while yielding signal intensities that are consistently higher (data not shown). In addition, the spot morphology on CMT-GAPS slides is much more uniform, with fewer "doughnuts" than on any of the alternatives we have investigated.

We also investigated a variety of spotting chemistries to determine which would provide the best results in subsequent hybridization assays. Figure 2 shows the results of a comparison between 50% dimethyl sulfoxide (DMSO) and 3× SSC as a spotting buffer. We have found that PCR products printed on CMT-GAPS aminosilane-coated glass microscope slides using 50% DMSO as a printing buffer result in the greatest hybridization intensities. This printing solution offers additional advantages. DMSO denatures the DNA, allowing better binding to the slide and providing more single-stranded targets for hybridization. Also, DMSO is hygroscopic and has a low vapor pressure, which allows DNA prepared for arrays to be stored for long periods of time without significant evaporation.

The print head on our arrayer and most others use "quill" pens that use capillary action to draw fluid into the spotting

Table 3. Aminosilane-Coated Slide Suppliers

Amersham Pharmacia Biotech	http://www.apbiotech.com
Corning	http://www.cmt.corning.com
Telechem	http://www.arrayit.com/

pens and surface tension interactions to dispense solution onto the slide. The Arrayit ChipMaker3™ microspotting pins (TeleChem International, Sunnyvale, CA, USA) are quite durable and can reproducibly generate high-quality spots with good precision; all array images shown were printed with the same set of ChipMaker3 pins over more than six months. Some parameters, such as the robot-arm acceleration, temperature and humidity, control spot morphology and size. We have found printing to be optimal at approximately 45% relative humidity and a constant temperature of 22°C. Changes in humidity and temperature have a significant impact on the size and morphology of spots and the efficiency of DNA binding to the slides. These parameters must be carefully controlled to provide consistent spotting. Figure 3 shows the effects of varying humidity and temperature on spot morphology and DNA retention.

DNA samples were spotted on the slides as described above, and temperature and humidity levels were recorded on a chart recorder. During the printing, temperature and humidity levels were allowed to vary continuously from 22.2°C and 45%–50% to a low of 16.7°C and 40%–45% and a high of 26.7°C and 80%–85% respectively. After hybridization with a vector-specific probe, we reconstructed the optimal printing
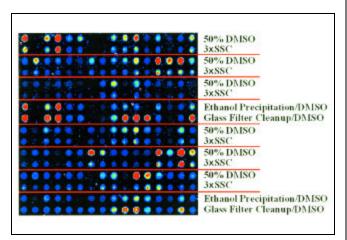

Figure 2. Effects of various spotting buffers and DNA purification protocols on DNA binding and hybridization using CMT-GAPS aminosilane-coated slides. This false-color image was generated by spotting identical samples in adjacent rows and hybridizing with a labeled mRNA probe; red lines separate paired rows. Paired rows 1–3 and 5–7 contain samples spotted using either 50% DMSO or 3× SSC as a spotting buffer. Comparing spots vertically adjacent to each other, it is clear that spotting with DMSO allows hybridization with significantly higher affinity than does spotting with SSC. In our evaluation, DMSO consistently gives 1.5-fold or greater hybridization intensities with 10% fewer drop-out spots. Paired rows 4 and 8 show the effects of different clean-up protocols on DNA binding and hybridization. The glass-filter method described in the text gives visibly better results than does simple ethanol precipitation.

BioFeature

conditions by using the chart recorder data to assign temperature and humidity values to the spots.

Arraying

1. Equal volumes of purified PCR product are added to DMSO in a 96-well, V-bottom plate (Corning). Typically, 5 µL each are used to prepare spotting plates that can be used to print 100 or more slides.

2. Slides to be printed are marked with a diamond-tipped pen; dust is removed by blowing the slides with high-pressure nitrogen gas, and the slides are placed in the arrayer. Care is taken not to touch the surface of the slides as oils adversely affect the ability of the slide surface to bind DNA.

3. Microtiter spotting plates are loaded into the arrayer and PCR products are spotted onto the slides at 22.2°C and 45% relative humidity.

4. After printing, the slides are allowed to dry and spotted DNA is bound to slide by UV-crosslinking at 90 mJ using a Stratalinker® (Stratagene, La Jolla, CA, USA).

5. Printed slides are stored in a light-tight box in a benchtop dessicator at room temperature until they are used for hybridization.
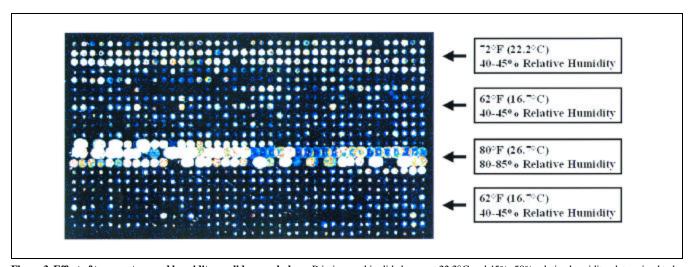
PROBE PREPARATION AND HYBRIDIZATION

Microarrays are used to assay differential gene expression by co-hybridization of fluorescently labeled probes prepared from different RNA sources. Similar to many other RNA-based assays, the purity and quality of the starting RNA have a significant effect on the subsequent results of the assay. Also, the products of the labeling reactions must be purified to remove unincorporated, labeled nucleotides that can produce a significant background on the slides following hy-

bridization. Finally, hybridization and washing conditions must be optimized to provide high specificity and minimize cross-hybridization. We have developed probe preparation and hybridization protocols using RNA derived from human carcinoma cell lines as a model system; variations of this protocol have been applied to the study of expression in rat and other systems.

RNA Extraction

Impurities in RNA preparations can have an adverse effect both on labeling efficiency and the stability of the fluorescent labels that are used for microarray expression analysis. We have found that TRIzOL® (Life Technologies) gives consistently high-quality RNA from cell culture and many tissue samples, although additional steps must be taken to remove polysaccharides when extracting RNA from some tissues. TRIzOL extraction is quick and produces a high yield of total RNA, as when used in the following protocol:

1. Media is aspirated from the cells and washed once with PBS.

2. Five milliliters of PBS are added and cells are scraped from the plate.

3. The cell suspension is transferred to a 50-mL polypropylene conical-bottom tube (Falcon).

4. The plate is washed with an additional 1 mL PBS, and the suspension is added to the tube.

5. The cells are centrifuged at $900 \times g$ for 3 min at 4°C, and the supernatant is discarded.

6. Two milliliters of TRIzOL/2×10^6 cells (approximately one 150-mm plate of fibroblasts) are added to the pellet, and the suspension is passed through an 18-gauge syringe several times to disrupt the pellet.

7. The sample is incubated at room temperature for 5 min.

Figure 3. Effect of temperature and humidity on slide morphology. Printing on this slide began at 22.2°C and 45%–50% relative humidity, determined to be optimal. Under these conditions, the spots have a uniform appearance. As printing progressed, the temperature was reduced to about 16.7°C, resulting in smaller, less distinct spots. Temperature and humidity were then increased. The rows of large spots in the center of the slide (rows 13–15) were printed at 26.7°C and 80%–85% relative humidity. As the temperature and humidity were decreased again, optimal conditions were achieved. By correlating data from a chart recorder with the spot number, we were able to determine the conditions that subsequently gave the best hybridization performance. Representative temperature and humidity levels are shown. The cDNA clones in this array were hybridized with a Cy3-labeled, vector-specific probe.

BioFeature

8. Chloroform, 0.4 mL (0.2 mL/1 mL TRIZOL) is added, shaken vigorously for 1 min and then incubated at room temperature for 2.5 min.

9. Cellular debris is removed by centrifugation at 2700× *g* for 15 min at 4°C.

10. The supernatant is transferred to 1.2-mL microcentrifuge tubes (0.5 mL/tube), and an equal volume of isopropanol is added to precipitate the RNA after incubation at room temperature for 15 min.

11. The mixture is centrifuged at 21 000× *g* for 15 min to pellet the RNA.

12. The supernatant is discarded, and the pellet is resuspended in 70% ethanol. The RNA can be stored in 70% ethanol at -20°C until used.

13. Before its use, the RNA is centrifuged at 21 000× *g* for 15 min at 4°C and the supernatant discarded.

14. The pellet is resuspended in diethylpyrocarbonate (DEPC)-treated water or in RNase-free TE buffer for labeling.

RNA Labeling

Because the ability to label small quantities of starting material is important for the study of expression in rare patient samples, we have focused on decreasing the quantity of starting material required. Probes for microarray analysis are prepared from RNA templates by the incorporation of fluorescently labeled deoxyribonucleotides during first-strand cDNA synthesis. Either total or poly(A$^+$) RNA can be used in the reverse transcription reaction. The oligo(dT) labeling of total RNA provides consistently higher quality probes from smaller quantities of starting RNA and without the added expense of poly(A$^+$) purification. Figure 4 shows the results of microarray hybridizations using labeled total or poly(A$^+$) RNA prepared from the same cell lines. An analysis of the fluorescence intensities for the elements in arrays hybridized with probes prepared from 1.5 µg poly(A$^+$) RNA (the equivalent 50–100 µg starting total RNA) and 4 µg total RNA indicate that total RNA labeling provides comparable probe activity without any increase in background fluorescence.

Typically, we prepare labeled probes using Cy3- and Cy5-dUTP, although Cy-labeled dCTP (both from Amersham Pharmacia Biotech, Piscataway, NJ, USA) can be used with an appropriate change in the concentrations of unlabeled dNTPs in the reaction. We have investigated some reverse transcriptases, including AMV and MMLV, and have found that SUPERSCRIPT™ II (Life Technologies) generates probes with significantly greater activity (data not shown).

Note that both Cy3 and Cy5 are photosensitive and care should be taken to minimize exposure to light during the labeling, hybridization, washing and scanning processes. On receipt, Cy-labeled nucleotides should be aliquotted into single-use, light-proof tubes and stored at -20°C until needed. All reactions should be carried out in foil-wrapped tubes and all hybridizations and washes in foil-wrapped containers.

Probe Labeling and Purification

1. A labeling-reaction master mixture should be prepared to contain 500 µM dCTP, 500 µM dATP, 500 µM dGTP, 100 µM dTTP, Cy3-dUTP/Cy5-dUTP, 400 U SUPERSCRIPT II, 1 mM dithiothreitol (DTT) and 1× reverse transcription buffer. Typically, we prepare sufficient quantities for 20 labeling reactions and store the unused solution at -20°C:

5× reverse transcription buffer	120 µL
DTT (5 mM)	60 µL
dATP (100 mM)	3 µL
dCTP (100 mM)	3 µL
dGTP (100 mM)	3 µL
dTTP (100 mM)	0.6 µL
DEPC-treated water	110.4 µL
Total	300 µL

2. To 10 µg total RNA [or 2 µg poly(A$^+$)] in a microcentrifuge tube, 2 µg oligo(dT) (18–20-mer; Life Technologies) and DEPC-treated distilled water are added to a total volume of 10 µL.

3. The reaction mixture is incubated at 70°C for 10 min and chilled on ice for 1 min.

4. To the RNA, the following are added:

Reverse transcription labeling mixture	15 µL
Cy3- or Cy5-dUTP (1 mM)	3 µL
SUPERSCRIPT II (200 U/µL)	2 µL

5. After being thoroughly mixed, the mixture is incubated at 42°C for 2 h.

6. The reaction is briefly centrifuged, and 1.5 µL 20 mM EDTA is added to stop the reaction.

7. Then, 1.5 µL 500 mM NaOH is added and heated at 70°C for 10 min to degrade the RNA.

8. The reaction is neutralized by adding 1.5 µL 500 mM HCl.

9. Unincorporated fluorescent nucleotides are removed by glass fiber filtration using GFX columns and the manufacturer's instructions (Amersham Pharmacia Biotech).

10. The purified products are eluted using 50 µL TE, pH 8.0, and the probe is completely dried in a SpeedVac® (Savant Instruments, Holbrook, NY, USA).

11. The probe is resuspended in 10 µL DEPC-treated water.

Hybridization

The goal in any hybridization is to obtain high specificity while minimizing background. We have developed protocols that give reproducible, high-quality hybridization results while maximizing the measured fluorescence from the array.

Aminosilane-coated slides bind DNA with high efficiency. Before hybridization, the free amine groups on the slide must be blocked or inactivated. Otherwise, nonspecific binding of labeled cDNA to the slide can deplete the probe and produce high background. Although the slides can be blocked chemically, we have found a simple prehybridization in a solution containing 1% BSA to be extremely effective in eliminating nonspecific binding of the probe to the slide.

Prehybridization has the advantage of washing unbound DNA from the slide before the addition of the probe. Any DNA that washes from the surface during hybridization

BioFeature

competes with DNA bound to the slide. As the kinetics of solution hybridization is much more favorable than surface hybridization, this can dramatically decrease the measured fluorescence signal from the microarray. All prehybridization and hybridization washes are carried out in microscope slide staining trays (VWR Scientific).

Prehybridization

1. Prehybridization buffer containing 5× SSC, 0.1% SDS and 1% BSA (Sigma, St. Louis, MO, USA) is prepared, followed by a 2× hybridization buffer containing 50% formamide, 10× SSC and 0.2% SDS.

2. Slides to be analyzed are placed in a Coplin jar (VWR Scientific); prehybridization buffer is added, and incubation is carried out at 42°C for 45 min.

3. The slides are washed by dipping five times in room-temperature Milli-Q-treated water.

4. The slides are then dipped in room-temperature isopropanol and air dried.

The slides should be used immediately after prehybridization. We have found that hybridization efficiency decreases rapidly if the slides are allowed to dry for more than 1 h.

Hybridization

1. Ten microliters each of purified Cy3- and Cy5-labeled probes are combined, mixed well and the following are then added to block nonspecific hybridization:

COT1-DNA (20 µg/µL)   1 µL   (Life Technologies)

poly(A)-DNA (20 µg/µL)   1 µL   (Amersham Pharmacia Biotech)

2. The probe mixture is heated at 95°C for 3 min to denature and centrifuged in a microcentrifuge set at maximum angular velocity for 1 min.

Table 4. Manufacturers of Microarray Scanners

Axon	http://www.axon.com
Beecher Instruments	
	http://www.beecherinstruments.com
GSI Lumonics	
	http://www.gsilumonics.com
Genetic Microsystems	
	http://www.geneticmicro.com
Genomic Solutions	
	http://www.genomicsolutions.com/
Molecular Dynamics	
	http://www.mdyn.com
Virtek	http://www.virtek.ca/

3. The probe is combined with an equal volume of 2× hybridization buffer that has been heated to 42°C.

4. The labeled probe is then applied to a prehybridized microarray slide and covered with a 22×60 mm polyethylene hydrophobic coverslip (PGC Scientific, Frederick, MD, USA).

5. The slide is placed in a sealed hybridization chamber (Corning), and 20 µL water is added to the chamber at the end of the slide.

6. The sealed chamber is placed in a 42°C water bath and incubated for 16–20 h.

7. The array is removed from the hybridization chamber, with care taken not to disturb the coverslip.

8. The slide is placed in a staining dish containing low-stringency wash buffer containing 1× SSC and 0.2% SDS at 42°C.

9. The coverslip is gently removed while the slide is in solution and agitated for 4 min.

10. The slides are washed at high stringency in a staining dish containing 0.1× SSC and 0.2% SDS at room temperature and agitated for 4 min.

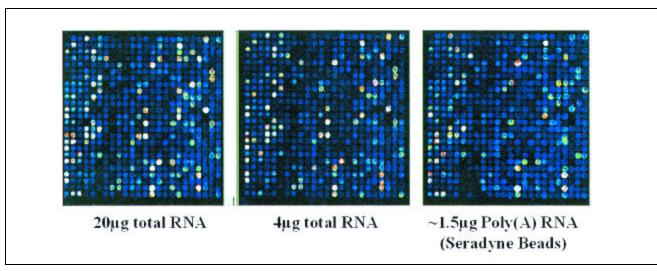20µg total RNA   4µg total RNA   ~1.5µg Poly(A) RNA (Seradyne Beads)

Figure 4. Comparison of hybridization probes made using total or poly(A⁺) RNA. Corresponding areas of cDNA microarrays containing 7200 elements, each hybridized with labeled probe prepared from 20 µg total RNA, 4 µg total RNA or 1.5 µg poly(A⁺) selected RNA. Note that the relative hybridization intensities are similar for each of the total RNA samples, and both are slightly greater than for the poly(A⁺)-hybridized sample. In our experience, 4–10 µg total RNA gives consistently high-quality hybridization results. All images were obtained using the same laser and PMT settings during scanning and are displayed using the same parameters.

Table 5. Microarray Image Processing Software Sources

BioDiscovery	http://www.biodiscovery.com/
Imaging Research	http://imaging.brocku.ca/Arrayvision.html
National Human Genome Research Institute	http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html
Stanford University	http://rana.Stanford.EDU/software/
The Institute for Genomic Research	http://www.tigr.org/softlab/

11. The slide is then washed in 0.1× SSC, agitated for 4 min and allowed to air dry.

DATA COLLECTION, NORMALIZATION AND ANALYSIS

Differential gene expression is assessed by scanning the hybridized arrays with a confocal laser scanner that is capable of differentiating both the Cy3- and Cy5-labeled probes and producing separate TIFF images for each. Similar to arraying robots, there are manufacturers that produce scanners capable of detecting Cy3 and Cy5 (Table 4), and most are planning to release instruments capable of detecting additional dyes.

Slide Scanning

We are currently using the ScanArray® 3000 (GSI Lumonics, Watertown, MA, USA). This scanner uses red and green helium-neon lasers operating at 633 nm and 543 nm to excite Cy5 and Cy3, respectively. Hybridized slides are scanned first in the Cy5 channel, and then the Cy3 channel, as Cy5 is more susceptible to photodegradation than Cy3. Data from each fluorescence channel is collected and stored as a separate 16-bit TIFF image. These images are analyzed to calculate the relative expression levels of each gene and to identify differentially expressed genes. The analysis process can be divided into two steps, image processing and data analysis. Figure 5 shows a typical hybridization image produced when things

BioFeature

Table 6. Microarray Data Analysis Sources

BioDiscovery	http://www.biodiscovery.com/
Silicon Genetics	http://www.sigenetics.com/
Spotfire	http://www.spotfire.com/
Stanford University	http://rana.Stanford.EDU/software/
TIGR	http://www.tigr.org/softlab/

work well. The contrast in this image has been adjusted to allow faint spots to be easily visualized. Important aspects of the hybridization are the low-level, uniform background and the good signal-to-noise ratio.

Image Processing

Image processing involves three stages. First, the spots representing the arrayed genes must be identified and distinguished from spurious signals that can arise from a precipitated probe, other hybridization artifacts or contaminants such as dust on the surface of the slide. This task is simplified to a certain extent because the robotic arraying systems produce a regular arrangement of the spotted DNA fragments. However, variable intensities, uneven slide backgrounds and some irregularities in the gridded arrays complicate the problem slightly. Generally, the problem of grid spot location is coupled with an estimation of the fluorescence background. For microarrays, it is important that the background be calculated
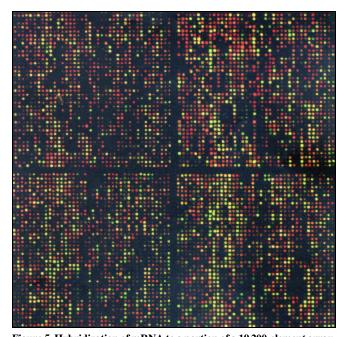

Figure 5. Hybridization of mRNA to a portion of a 19 200-element array. Using the protocols described here, 10 µg RNA extracted from related human colon carcinoma test and reference cell lines (KM12L4A and KM12C, respectively) were reverse transcribed and labeled with Cy5-dUTP and Cy3-dUTP, respectively. These were then hybridized to a microarray containing 19 200 distinct human cDNA clones. The contrast on the image has been adjusted to allow the majority of the spots in the array to be easily visualized. The protocols outlined here consistently provide similar hybridization results.

locally for each spot rather than globally for the entire image because uneven background can often arise during the hybridization process. The second step in analysis of the array images is the estimation of background.

Following spot identification and local background determination, the background-subtracted hybridization intensities for each spot must be calculated. There are currently two schools of thought regarding the calculation of intensities: the use of the median and the mean intensity for each spot. As array analysis generally uses ratios of measured Cy3-Cy5 intensities to identify differentially expressed genes, the mean and the integrated intensities are operationally equivalent. In comparisons of intensities measured for normalization controls added to the labeling reactions, we have found mean intensities to give more consistent results. Consequently, we have used these in subsequent calculations.

Table 5 lists some image-processing software packages that are available. We have developed a software package called TIGR Spotfinder for image processing (http://www.tigr.org/softlab/). This software uses a thresholding algorithm that separates spots from the background and allows a grid to be laid across the spots. Having found a grid, spots are found within each grid element, the local background is calculated and background-subtracted; integrated intensities are then calculated in both the Cy3 and Cy5 channels. Measured intensities are entered into the Molecular Analysis of Gene Expression (MAGE), a relational database (Sybase, Emeryville, CA, USA) specifically designed to capture gene expression data.

Data Normalization and Analysis

Following image processing, the data generated for the arrayed genes must be further analyzed before differentially expressed genes can be identified. The first step is the normalization of the relative fluorescence intensities in each of the two scanned channels. This is necessary to adjust for differences in labeling and detection efficiencies for the fluorescent labels and for differences in the quantity of starting RNA from the two samples examined in the assay. These problems can cause a shift in the average ratio of Cy5 to Cy3, and the intensities must be rescaled before an experiment can be properly analyzed.

The normalization strategies are based on some underlying assumptions regarding the data; the strategies used for each experiment should be adjusted to reflect both the system under study and the experimental design. The primary assumption is that for either the entire collection of arrayed genes, some subset such as housekeeping genes or for some added set of controls, the ratio of measured expression averaged over the set should be 1.

Depending on the experimental design, there are three useful approaches for calculating normalization factors. The first simply uses total measured fluorescence intensity. The assumption underlying this approach is that the total mass of RNA labeled with either Cy3 or Cy5 is equal. While the intensity for any one spot may be higher in one channel than the other, these fluctuations should average out over thousands of spots in the array. Consequently, the total integrated intensity across all the spots in the array should be equal for both channels. Alternatively, one could add a number of controls in increasing, but the

BioFeature

equimolar concentrations to both the labeling reactions and the sum of the intensities for these spots should be equal.

A second approach uses linear regression analysis. For closely related samples, one would expect many of the genes to be expressed at nearly constant levels. Consequently, a scatterplot of the measured Cy5 versus Cy3 intensities should have a slope of 1. Measured intensities for added equimolar controls should behave similarly. With this assumption, one can use regression analysis techniques to calculate the slope, which is then used to rescale the data and adjust the slope to 1.

A third approach (2) relies on the assumption that some subset of housekeeping genes exists and that for these, the distribution of transcription levels should have some mean value and SD, independent of the sample. Here, the ratio of measured Cy5 to Cy3 ratios for these genes can be modeled and the mean of the ratio adjusted to 1. Chen and collaborators (2) describe an iterative procedure to achieve this normalization. We have implemented their algorithm, a variation of it that uses the entire data set and the total intensity and linear regression normalization into a data visualization and analysis tool called TIGR ArrayViewer. The ArrayViewer is free and can be obtained from http://www.tigr.org/softlab/.

In any normalization approach, care must be taken in handling genes expressed at low levels. Statistical fluctuations in the measured levels can cause a significant variation in the ratios that are calculated, and inefficiencies in labeling for either of the two dyes can cause low-intensity genes to disappear from the arrays. Typically, we only use spots in the final analysis where the intensities in both channels are two standard deviations above background.

Following normalization, data are analyzed to identify genes that are differentially expressed. Most published studies have used a post-normalization cutoff of two-fold up- or down-regulation to define differential expression. The approach defined by Chen et al. (2) provides confidence intervals that can be used to identify differentially expressed genes. To separate genes that are truly differentially expressed from stochastic changes, we typically conduct three independent microarray assays, starting from independent mRNA isolations, and define differential expression based on their consensus.

There are a growing number of programs for image processing and data analysis that are available from academic and commercial organizations. Tables 5 and 6 include a number of widely used sources.

CONCLUSION

The examination of gene expression using microarrays holds great promise for the identification of candidate genes involved in a variety of processes. Indeed, the experiments that have been described to date have confirmed known patterns of expression and have provided information on genes of unknown function. However, most applications have only allowed the identification of genes differentially expressed at significant levels. The true challenge—and the promise of this technique—will be to use it to identify genes that are consistently up- or down-regulated by 10% or 20%, yet play significant roles in the development and progression of disease.

This will require the analysis of data from multiple experiments and the correlation of patterns of gene expression with additional experimental and clinical information.

Recently, hierarchical clustering (4) and self-organizing maps (11) have been applied to the analysis of microarray expression data across multiple experiments. However, each of these depends on having reliable and reproducible data from each microarray assay. The laboratory techniques outlined here have allowed reproducible hybridization results (Figure 5). Although these protocols will continue to evolve, we believe they represent a reliable starting point for those beginning microarray experimentation.

REFERENCES

1. **Boguski, M.S. and G.D. Schuler.** 1995. ESTablishing a human transcript map. Nat. Genet. *10*:369-371.
2. **Chen, Y., E.R. Dougherty and M.L. Bittner.** 1997. Ratio-based decisions and the quantitative analysis of cDNA microarray images. J. Biomed. Optics *24*:364-374
3. **Eisen, M.B. and P.O. Brown.** 1999. DNA arrays for analysis of gene expression. Methods Enzymol. *303*:179-205.
4. **Eisen, M.B., P.T. Spellman, P.O. Brown and D. Botstein.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*:14863-14868.
5. **Liang, P. and A.B. Pardee.** 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science *257*:967-971.
6. **Lockhart, D.J., H. Dong, M.C. Byrne, M.T. Follettie, M.V. Gallo, M.S. Chee, M. Mittmann, C. Wang et al.** 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotechnol. *14*:1675-1680.
7. **Quackenbush, J., F. Liang, I. Holt, G. Pertea and J. Upton.** 2000. The TIGR gene indices: reconstruction and representation of expressed gene sequences. Nucleic Acids Res. *28*:141-145.
8. **Schena, M., D. Shalon, R.W. Davis and P.O. Brown.** 1995. Quantitative monitoring of gene expression patterns with complementary DNA microarray. Science *270*:467-470.
9. **Schuler, G.D.** 1997. Pieces of the puzzle: expressed sequence tags and the catalog of human genes. J. Mol. Med. *75*:694-698.
10. **Shalon, D., S.J. Smith and P.O. Brown.** 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. *6*:639-645.
11. **Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander and T.R. Golub.** 1999. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA *96*:2907-2912.
12. **Velculescu, V.E., L. Zhang, B. Vogelstein and K.W. Kinzler.** 1995. Serial analysis of gene expression. Science *270*:484-487.

Address correspondence to:

Dr. John Quackenbush
The Institute for Genomic Research
9712 Medical Center Drive
Rockville, MD 20850, USA
e-mail: johnq@tigr.org