

## **cDNA Microarrays in Neuroscience: nylon membrane based arrays**

<sup>1</sup>K. G. Becker, <sup>1</sup>T. Barrett <sup>2</sup>M. P. Vawter, <sup>1</sup>W. H. Wood III, and <sup>1</sup>C. Cheadle

<sup>1</sup>DNA Array Unit, National Institute on Aging, <sup>2</sup>Cellular Neurobiology Branch, National Institute on Drug Abuse; National Institute of Health, Baltimore, MD 21224

Address correspondence to:

K. G. Becker Ph.D.  
DNA Array Unit, Room 4D16, GRC,  
National Institute on Aging,  
5600 Nathan Shock Drive,  
National Institute of Health, Baltimore, MD 21224

Tel: 410-558-8360

Fax: 410-558-8281

beckerk@grc.nia.nih.gov

[www.grc.nia.nih.gov/branches/rrb/dna](http://www.grc.nia.nih.gov/branches/rrb/dna).

## 1. Introduction

cDNA arrays allow the simultaneous analysis of gene expression patterns of thousands of genes under a similar set of experimental conditions, thus making the data highly comparable (Schena 95, Duggan 99). In some cases arrays are used simply as a primary screen leading to downstream molecular characterization of individual gene candidates. In other cases, the goal of expression profiling is to begin to identify complex regulatory networks underlying developmental processes and disease states. cDNA arrays were originally used with cell lines (Derisi 96) or other simple model systems (Schena 95). More recently, cDNA arrays have begun to be used in the analysis of more complex biological systems including the brain (Whitney 99, Tanaka 00, Luo 99, Colantuoni 00).

The application of cDNA arrays in Neuroscience has lagged behind other fields for a number of reasons. These include a requirement for a large amount of input probe RNA in fluorescent-glass based array systems and the cellular complexity introduced by multicellular brain and neural tissues. An additional factor that impacts the general use of arrays in neuroscience is the lack of availability of sequenced clone sets from model systems. While human cDNA clones have been widely available, high quality rat, mouse, and drosophila, among others are just becoming widely available. A final factor in the application of cDNA arrays in Neuroscience is cost of commercial arrays. The use of cDNA arrays in increasingly complex biological systems requires multiple repetitions to increase statistical significance and experimental confidence. This drives up the cost per experiment using commercial arrays in complex systems as compared to simple cell line based systems. As academic cDNA array facilities become more commonplace custom made arrays will become more widely available at a lower cost allowing more widespread applications.

There are generally three types of gene expression arrays; oligonucleotide chips, cDNA arrays printed on glass slides using fluorescent probes, and cDNA arrays printed on nylon filter membranes. Each format has associated advantages and disadvantages.

Oligonucleotide gene expression arrays, generally available from Affymetrix, Inc., will not be discussed here. cDNA PCR products printed on coated glass microscope slides in combination with fluorescent labeled cDNA probes have been used with great success in many different systems. Glass based array systems generally will not be discussed here except for comparative purposes.

Differential screening of cDNA libraries using filter membranes with radioactively labeled complex probes has been performed successfully since 1983 (Sargent 83). This approach has historically used a single radioactive label comparing the hybridization pattern on multiple membranes, this is in contrast to the simultaneous two-color hybridization of the glass-fluorescent approach (Schena 95).

Four specific advantages of glass/ fluorescent based array systems as compared to membrane/radioactive based systems are 1) very close packing distance between spots 2) the ability to perform simultaneous two-color hybridization, 3) low-hybridization volumes which concentrate the labeled probe, and 4) the use of non-radioactive labels in the context of safety and regulatory concerns.

Membrane/ radioactive systems however offer multiple advantages as compared to glass/fluorescent based systems as well;

- (1) Higher efficiency of labeling cDNA with radioactive labels as compared to fluor labels which result in higher specific activity probes. This translates into using 10 to 100 fold less total RNA per labeling (1-10 ug with 33-P vs 50-200 ug Cy-3, Cy5). In cases where RNA is limiting, this allows a greater chance for multiple replications leading to statistical significance (Bertucci 99).

- (2) A single label system as compared to a dual label system. This allows direct comparison of two cDNA probes. While not only having lower nucleotide incorporation rates than <sup>33</sup>P-dCTP, Cy-3 and Cy-5-d-UTP also have different incorporation rates as compared to each other, with Cy-5-d-UTP being considerably less efficient than Cy-3. This can lead to difficult comparisons between two fluorescent probes labeled at unequal efficiencies.
- (3) Multiple use of nylon filters. Filter membranes can be used three to five times. This reduces cost, and allows direct comparisons on the same array. Significant reduction of costs allows replication of hybridizations.
- (4) Universal protocols and accessibility. Nylon membrane based technology, including radioactive labeling, is widely used and widely accessible using existing reagents, hardware and software. In many cases may result in lower setup and training costs, and shorter time in generating data.

Overall, the choice to use nylon membranes as a cDNA support combined with radioactive probes influences many aspects of the array process as compared to the glass slide-fluorescent probe approach. These include; producing membrane based cDNA arrays, experimental strategies that vary from the fluorescent approach, data acquisition and image processing, quantitative analysis, normalization of multiple arrays using a single probe system, statistical analysis, biological analysis, data visualization, and data presentation. These aspects of the production, use, and analysis of cDNA arrays on nylon membrane supports using radioactive labels as performed in our laboratory will be discussed here. This is not intended to be a survey of the entire field of hybridization of nylon membranes. It should allow the interested investigator to get an overview of the entire process in our laboratory and provide resources to get additional details concerning each step.

**\*A note on nomenclature.** In this paper the terms probe and target are used as commonly used in molecular biology. Probe refers to the labeled unknown pool of cDNA or RNA while target(s) refers to the known cDNA spot(s) attached to a solid support

## **2. Printing nylon membrane cDNA arrays using an Affymetrix 417 arrayer.**

Printing cDNA arrays on Nylon membranes with the Affymetrix 417 arrayer is rapid and efficient. The production of glass arrays can take up to one month which may include coating and aging of slides, printing, further aging, denaturation and chemical processing

Arraying resuspended PCR products onto nylon membranes takes approximately 1 day with membrane preparation, arraying, cross-linking and baking.

### *cDNA clone Preparation*

The preparation of cDNA products for printing on nylon membranes is essentially identical to preparation for printing onto glass slides (<http://www.wenet.net/~telechem/DNA-Microarray-Protocols/>), with the major exception of the addition of NaOH. Briefly, purified plasmids containing cDNA clones are PCR amplified in 96-well plate format in 100 ul PCR reactions using library specific primers. Two microliters of each PCR reaction is tested for single band purity, and for relative quantity by gel electrophoresis. The remaining PCR reactions are transferred to 96-well V bottom plates (Corning, Inc. cat # 3894), ETOH precipitated and air dried overnight. These purified pellets are resuspended in 40 ul of 1X TE. Immediately prior to arraying, NaOH is added to 0.1 N. The final concentration of denatured PCR product is approximately 100 ug/ ml.

### *Membrane preparation*

The 417 arrayer has six sections on the printing platen that accommodate 7 glass slides each for a total capacity of 42 glass slides. Six stainless steel forms were fabricated to fit onto the printing surface replacing the space that held the individual glass slides. These forms are approximately 160 mm x 80 mm x 3 mm each.

Nytran<sup>+</sup> Supercharge (S&S, Inc., cat # 10416296) nylon membrane is cut to fit each form exactly keeping the protective red paper covering on top of the filter membranes. Krylon repositionable adhesive (cat # 7020) is then sprayed lightly onto each form. The glue is allowed to dry a few minutes. Too much glue can cause the reverse side of the nylon membrane to pull off when the membranes are removed from the forms after printing. Too little glue and the membrane can come free and get twisted in the pin/ring head causing damage to the print head of the arrayer. The cut nylon membranes are then placed on a clean surface and the form is placed on-top with the glue side down. Even pressure is applied to the entire form. The edges of the membrane are trimmed with a razor. Gloves must be worn at all times.

The form/membrane/paper units are placed into the arrayer slots and pushed to the left to engage the springs and pushed forward/up until the edge is even with the top of the slot. This process is repeated to fill the entire platen with form/membrane units. This allows 42 microscope slide size areas to be printed.

### *Arraying*

The pin head should be calibrated to the top surface of the form/membrane units in a similar manner as 417 arrayer calibration onto glass slides.

Addition of a volume of NaOH to give a final concentration of 0.1 N NaOH is done just before the first printing of each set of three plates (4.4 ul of 1 N NaOH per well) and lightly vortexed. The first 3 plates are then loaded into the arrayer. Plates that have been printed once will not need NaOH but may need to have 5 to 10 ul of H<sub>2</sub>O added to correct for evaporation.

For use with radioactive probes, we spot using a 300 micron pin, with spot spacing of 665 microns center to center. This allows 2304 individual spots on the printable area the size of a 25 x 75 mm microscope slide. This produces 16 identical 12 x 12 grids, or subarrays, of 2304 individual spots from 24 96-well plates. In many cases we print 12 plates in duplicate on this same area. That allows 1152 spots to be arrayed in duplicate. This takes approximately 8 hours of continuous printing to produce 42 identical arrays. After each set of three plates, the 96-well plates are removed from the arrayer, sealed with plastic plate sealers, and returned to the -20 freezer for long term storage.

### *Post-arraying processing*

The form/membrane units are removed from the arrayer one at a time and numbered in the upper right hand corner with their print position 1-6. The entire units are UV crosslinked twice using a Stratalinker 2400 (Clonetech, Inc.) at 120 millijoules/ sec.

At this time, the membranes should be quality controlled by sight under a dissecting microscope to note any missing spots. Any position without a dent in the membrane means the sample was not picked up by the ring, and no DNA deposited. It appears that printing with a pin with no liquid does not produce an indentation in the nylon membrane and can be easily seen. These positions should be marked on the array map to store with the filters. Membranes are then carefully removed from the supports and placed between pieces of blotting paper and baked in the 70-degree oven for between 1 and 2 hours. Each individual filter from 1-42 is then labeled with the name, date and number using a 00/0.3 width Koh-I-Noor rapidograph and rapiddraw 3084-F ink, cut apart with a razor blade and stored between blotting paper at room temperature until use. The stainless steel arraying forms are then cleaned with 100 % ETOH and air dried. These arrays can be used immediately. We

regularly, array, cross-link, and bake a set of 2304 clones in ~10 hours, then use them directly to pre-hybridize and hybridize overnight.

Each round of printing produces 42 identical arrays and consumes approximately 4  $\mu$ l of product for each arraying run. PCR products in 96-well V-bottom plates initially in ~44  $\mu$ l of liquid routinely allow the printing of approximately 200-250 individual arrays. The number of arrays produced from one PCR preparation depends greatly upon evaporation of the PCR products during storage, printing, and processing. Efforts to increase humidity during dry seasonal periods can increase the total number of arrays produced. A detailed protocol for arraying on nylon membranes with an Affymetrix 417 arrayer can be found at the following WWW address: <http://www.grc.nia.nih.gov/branches/rrb/dna/dna.htm>.

### 3. cDNA probe labeling and membrane hybridization protocols.

These protocols are relatively standard in the world of molecular biology for both cDNA labeling and filter hybridization. These work for us with high efficiency and are universally available. There are probably better similar protocols available for both labeling and hybridization. There are commercial hybridization solutions that may increase hybridization signals and lower hybridization time, as well as many different labeling strategies that work well with 33-P dUTP.

#### *Radioactive cDNA probe preparation:*

cDNA probes are prepared as previously described (Whitney 99). Briefly, 3-10  $\mu$ g of total RNA is mixed with 1  $\mu$ l of 1  $\mu$ g/ $\mu$ l 10-20-mer poly(dT) primer (Research Genetics Cat # Poly T.GF). H<sub>2</sub>O is added to 15  $\mu$ l. The mixture is heated at 65°C for 5 min, followed by incubation on ice for 2 min.

The reverse transcription reaction mixture is then added:

8  $\mu$ l of 5 X first strand PCR buffer (LTI, Inc. Cat # 18084-014),  
 4  $\mu$ l of 20 mM dNTPs (-dCTP) (Pharmacia Cat # 27-20X0-0),  
 4  $\mu$ l of 0.1 M DTT (LTI, Inc Cat # 18084-014),  
 1  $\mu$ l (40 U) of RNaseOUT (LTI Cat # 10777-019),  
 5  $\mu$ l of 3000 Ci/mmol  $\alpha$ -<sup>33</sup>P dCTP (ICN Cat # 58430)  
 H<sub>2</sub>O to a final volume of 38  $\mu$ l.

Two  $\mu$ l of Superscript II reverse transcriptase (LTI, Inc Cat # 18084-014) is then added and the sample is incubated for 35 min at 42°C followed by additional 2  $\mu$ l of Superscript II reverse transcriptase and another 35 min incubation at 42°C.

Five  $\mu$ l of 0.5 M EDTA is then added to chelate divalent cations. After addition of 10  $\mu$ l of 0.1 M NaOH, the samples are incubated at 65°C for 30 min to hydrolyze the template RNA. Following the addition of 25  $\mu$ l of 1 M Tris, pH 8.0, the samples are purified using Bio-Rad 6 purification columns (Bio-Rad, Inc. Cat # 732-6223).

This results in a labeled product of approximately 75  $\mu$ l. One microliter is counted in a liquid scintillation counter. Typical labeling generally yields 3-5 x 10<sup>7</sup> total CPM for 75 microliters. An online protocol for cDNA probe preparation can be found at the following web address: [www.grc.nia.nih.gov/branches/rrb/dna](http://www.grc.nia.nih.gov/branches/rrb/dna).

For a protocol for labeling cDNA with 33-P-dCTP for use on glass slides:

<http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/protocol.html>

### Hybridization:

25 mm x 75 mm Nylon cDNA arrays are loaded into empty 50 ml disposable Falcon tubes (Falcon cat # 35-2098) with the DNA side pointed in. Up to three membranes may fit into one 50 ml tube for use with one probe. To pre-wash the filters, the tube is filled with 50 mls of 2X SSC, decanted and drained on a paper towel. This attaches the membrane to the side-wall of the tube. Any bubbles under the membrane are removed by rolling a 5 ml plastic pipette against the membrane.

cDNA microarrays are pre-hybridized in hybridization buffer containing 4.0 ml Microhyb (Research Genetics, Cat # Hyb250.GF), 10 ml of 1 mg/ml human Cot 1 DNA (denatured at 950C for 5 min prior to use; LTI, Inc., Cat # 15279-001) and 10 ml of 8 mg/ml poly(dA) (Sigma, Inc. Cat # P-9403), denatured at 950C for 5 min prior to use.

Pre-hybridization and hybridization is performed at 500C in a standard hybridization oven (HybAid, Inc. # HS 9360) for 4 hours. After 4 hours of pre-hybridization at 500C, the entire labeled probe (~ 75 ul, 3-5 x 10<sup>7</sup> CPM) is denatured for 5 min at 950C and added to the ~4 ml pre-hybridization solution. This is then followed by 12-18 hours of incubation at 50 0C.

After hybridization, the 4 ml solution is decanted. The hybridized arrays are washed quickly in the original tube with 50 ml of 2 X SSC and 1% SDS pre-heated to 50 0C. This is followed by 1-2 times of washing in 2 X SSC and 0.1% SDS at 500C for 15 min each. The arrays should be monitored with a survey meter at each wash step. The need for further washing should be determined empirically. If more stringent washing is needed, the filters can be washed in 1X SSC and 0,1% SDS at 650C for 5-15 minutes.

The wet membrane arrays are then aligned on a glass or metal plate. All excess fluid and micro-bubbles are removed. The filters are covered very tightly with plastic wrap, which is taped securely to the back of the plate. A small piece of moistened blotting paper may be placed at the bottom to keep the filters moist. If filters are to be re-used, do not let them dry out.

The microarrays are then exposed to phosphorimager screens for 1 to 3 days. The screens are then scanned in a Molecular Dynamics STORM PhosphorImager at 50 µm resolution.

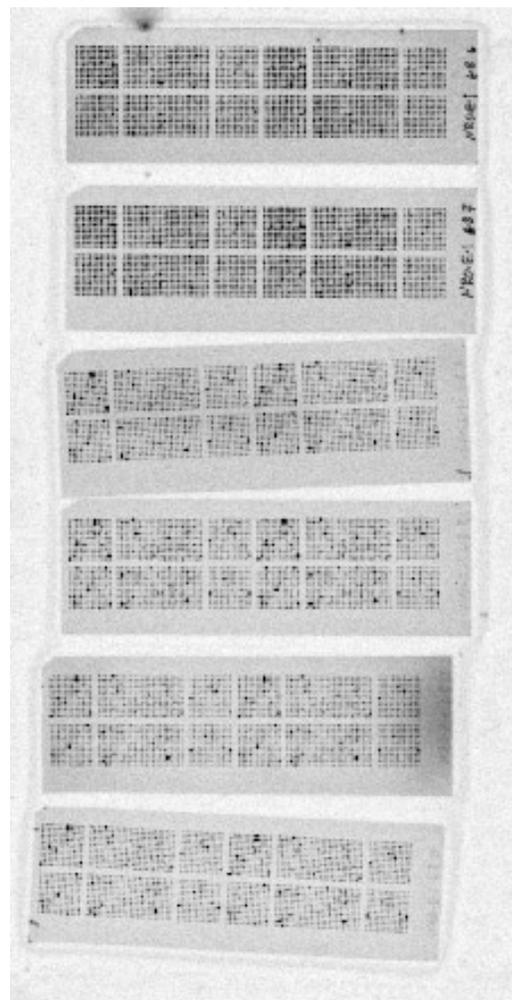
### Probe stripping and re-use

For stripping and re-use of the filters, wash 2 times at 65 0C in 0.4N NaOH/ 0.1%SDS for 30min each in ~200-300 mls solution w/ vigorous shaking. Then wash 2 times at room temp in 0.2 M Tris-HCL (p.h..8.0), 1xSSC and 0.1%SDS for 10min each. Air dry, expose to phosphor screen overnight to determine stripping efficiency.

### 4. Image Acquisition and Signal Quantitation

Radioactive hybridization can be visualized either by classical autoradiography followed by digital image scanning or by the more recent phosphorimager technologies. We use a Molecular Dynamics Storm PhosphoImager (Molecular Dynamics, Inc.) scanning at 50 micron resolution. Other models of phosphorimagers are used as well. The resulting images, similar to the one shown in figure 1, are usually composites of multiple hybridization experiments. Digital enhancement

Fig 1



of the images is routinely used to provide the best conditions for visual inspection. However this image manipulation does not affect the basic signal calculations as discussed below.

The first task in analysis after acquiring the primary composite image is to separate each filter/hybridization into its own unique image file. We use the IQ Tools software (Molecular Dynamics, Inc.) for this process of cropping and saving each individual filter image separately. In addition, whole images are straightened along their horizontal axis using the IQ Tools image rotation function. This straightening process is required later in the downstream quantitation analysis procedures.

Fig 2

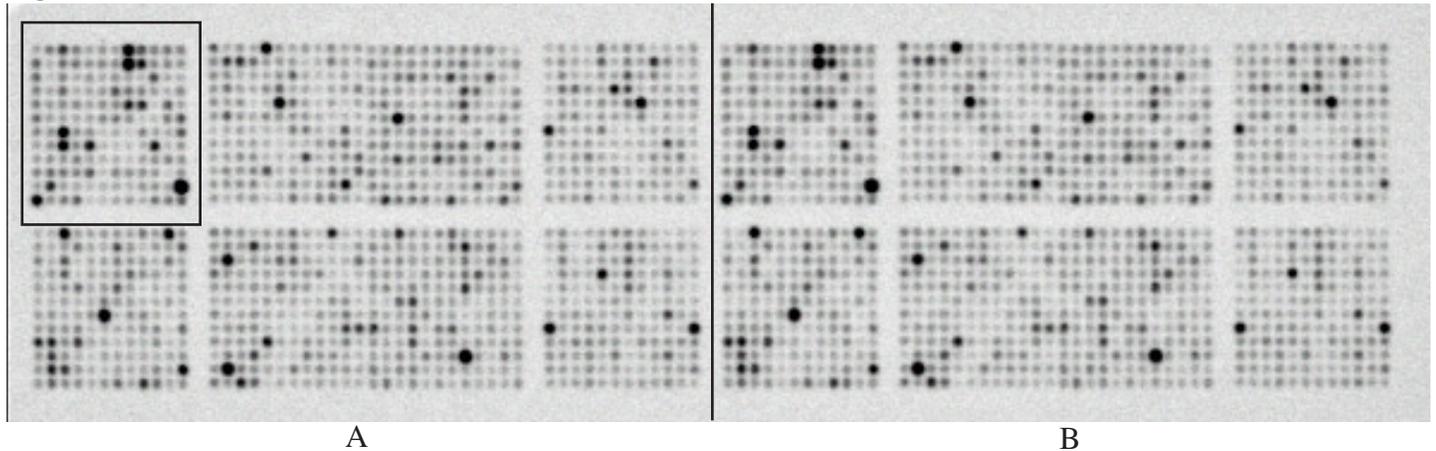
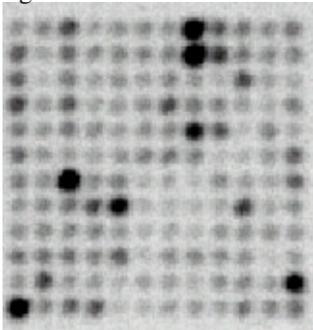


Figure 2 is an example of a cropped and straightened filter/hybridization image enlarged for closer visual inspection. The printing pattern of an entire array is composed of a variable number of subarrays, each of which is in turn composed of a 12X12 grid of 144 features. This particular array, the NIA Neuroarray is printed in duplicate left and right (A & B) (as indicated by the line). Each duplicate is composed of 8 individual subarrays, one of which is highlighted in the boxed area at the upper left hand corner. For simplicity, we will demonstrate the overlay and quantitation procedures using this highlighted subarray area.

Fig 3a



It is useful for display purposes to perform two color overlays between different filter/hybridization images. The image can often highlight major gene expression changes which will be validated in the quantitated data. Once both black and white images (Fig 3A& B) are aligned, a false color overlay (Fig 3C) is generated from grey scale images loaded into either red or green channels in programs such as Adobe Photoshop. The black and white images are then inverted to a black background with white dots. The third channel is then painted black for contrast.

Fig 3b

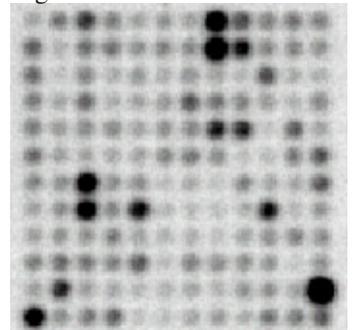
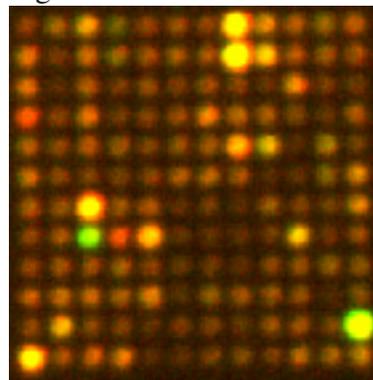


Fig 3c



—red channel

—green channel

—invert image

—black in 3rd channel

The green and red signals correspond to the presence or absence of signal in filters 3A or 3B.

### Image Quantitation

Image data is quantitated using ImageQuant (IQ) software (Molecular Dynamics, Inc.) on the original cropped and rotated image. A grid is placed over the image matrix (Fig 4) and all pixels within each grid square corresponding to the signal area of a single array element are counted on a 1-256 greyscale index. The resulting data is generated in the form of a "volume report" and is exported by the IQ program into an Excel spreadsheet format. This data then can be cut and pasted directly into a Excel spreadsheet template which translates the subarray grid position into the array element gene identity. A typical spreadsheet is shown below (Fig 5) The imported volume report data (highlighted in red) with the array grid position to its left and the gene identities to its right. This data is now ready for the next steps of normalization and quantitative gene difference calculations.

Fig 4

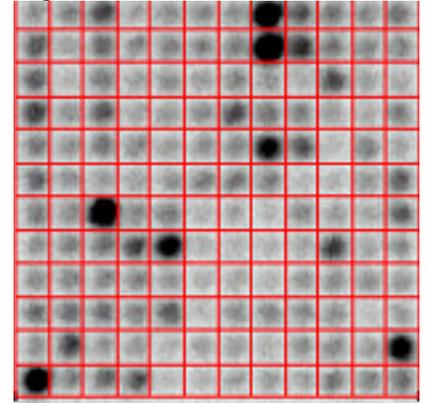


Fig 5. Spreadsheet with inserted pixel intensity values

Index	Name	aVMB	eVMB	GeneName	SA	R	C	RC	96WP	96WR	96WC	Pin #	T
1	GRID- 1-R1C1	5559.47	5913.54	ESTs	1	1	1	A12	2	A	12	2	
2	GRID- 1-R2C1	9941.22	11703.93	ESTs,Highlysimilaroneurin R.norvegic	1	2	1	A10	2	A	10	2	
3	GRID- 1-R3C1	9509.49	11235.04	TYROSINE-PROTEINKINASERECEPTO	1	3	1	A8	2	A	8	2	
4	GRID- 1-R4C1	12525.70	8516.85	Enolase2,(gamma,neuronal)	1	4	1	A6	2	A	6	2	
5	GRID- 1-R5C1	7161.05	8789.04	GTPcyclohydrolase1(dopa-responsivec	1	5	1	A4	2	A	4	2	
6	GRID- 1-R6C1	8039.57	10056.91	Glialfibrillaryacidicprotein	1	6	1	A2	2	A	2	2	
7	GRID- 1-R7C1	6872.51	7754.47	Myelinbasicprotein	1	7	1	A2	1	A	12	2	
8	GRID- 1-R8C1	4769.61	5079.77	Nervegrowthfactorbeta	1	8	1	A10	1	A	10	2	
9	GRID- 1-R9C1	5472.06	5410.40	Opioidreceptor,kappa1	1	9	1	A8	1	A	8	2	
10	GRID- 1-R10C1	7114.85	7602.45	Protein kinaseC, alpha	1	10	1	A6	1	A	6	2	
11	GRID- 1-R11C1	6515.13	6195.17	Ric(Drosophila)-like(expressedinneuro	1	11	1	A4	1	A	4	2	
12	GRID- 1-R12C1	40489.30	40963.89	Tyrosine3-monooxygenase/tryptophan	1	12	1	A2	1	A	2	2	
13	GRID- 1-R1C2	6399.54	9232.75	MJD1	1	1	2	C12	2	C	12	2	
14	GRID- 1-R2C2	4417.25	4618.55	MonooamineoxidaseA	1	2	2	C10	2	C	10	2	
15	GRID- 1-R3C2	3380.64	3745.57	Nervegrowthfactorreceptor	1	3	2	C8	2	C	8	2	
140	GRID- 1-R7C12	9251.97	12912.00	cytoskeleton associated protein (CG22)	1	7	12	G12	5	G	12	2	
141	GRID- 1-R8C12	5111.12	5939.17	guanylate kinase (GUK 1)	1	8	12	G10	5	G	10	2	
142	GRID- 1-R9C12	4841.55	6829.13	tubulin-folding cofactor E	1	9	12	G8	5	G	8	2	
143	GRID- 1-R10C12	4314.13	7945.52	Integrin, alpha E (antigen CD103, mucos	1	10	12	G6	5	G	6	2	
144	GRID- 1-R11C12	26717.65	180086.82	Protein phosphatase 1, catalytic subuni	1	11	12	G4	5	G	4	2	
145	GRID- 1-R12C12	8337.51	14996.20	Protein phosphatase 2 (formerly 2A), re	1	12	12	G2	5	G	2	2	
Average		7700.42	10567.05										

## 5. Normalization of array data

Normalization of the data from multiple array hybridizations is necessary when non-specific background intensity is unequal. There are generally two different classes of non-specific background; 1) local variation across a single array and 2) uniformly higher background across the entire array as compared to other arrays. Local non-uniform background can appear in many forms; speckles, streaks, smears, etc. Generally the best solution to high local background is to re-do the hybridization. Normalization for local background will not be discussed here. Normalization strategies for cDNA arrays, including local background, have recently been described (Schuchhardt 00).

The second class of background requiring normalization is uniform differences in background. This can be due to differences in starting RNA quality, non-specific contaminants in the probe mixture, or to true global transcriptional differences in the two RNA populations. The strategies described below are suited for uniform differences in background.

Most normalization strategies work well with small differences between any two images. As the differences between any two images gets greater and more non-uniform, normalization strategies begin to break down and can distort the underlying data.

No single normalization approach can correct for all experimental variations in array images. Useful approaches used in our laboratory are described below.

### *Simple Normalization*

Simple normalization works reasonably well in many common cases of uniform background differences. This approach is not unlike simple normalization in other biological assays. Figure 6A and 6B shows with two images of unequal total intensity and uniform background. The data is quantitated from each image and a total intensity value is acquired for each image (Fig 6C). A normalization factor is determined by dividing the average total intensity by the true total intensity. Each original column is then multiplied by its appropriate normalization

Fig 6a

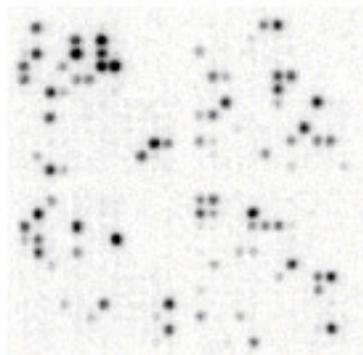


Fig 6b

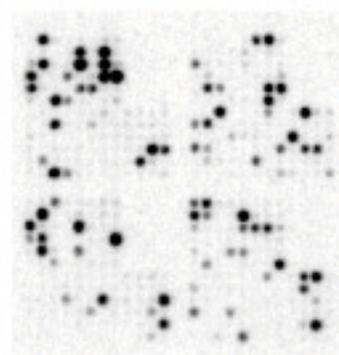


Fig 6c

<b>Raw Intensity Values</b>					
<b>Gene Name</b>	<b>Image A</b>	<b>Image B</b>	<b>Image An(2.03)</b>	<b>Image Bn(0.66)</b>	<b>Bn/An</b>
Gene 1	199	521	404	344	0.85
Gene 2	192	361	389	238	0.61
Gene 3	1000	3221	2030	2126	1.05
Gene 4	8863	27738	17992	18307	1.02
Gene 5	6952	21086	14113	13917	0.99
Gene 6	313	991	634	654	1.03
Gene 7	111	247	224	163	0.73
Gene 8	112	229	228	151	0.66
Gene 9	176	534	358	352	0.98
Gene 10	105	402	213	266	1.25
<b>Sum</b>	<b>18023</b>	<b>55328</b>			
<b>Ave</b>	<b>36675.5</b>				
	<b>Ave/SumA=2.03</b>				
	<b>Ave/SumB=0.66</b>				

factor. Ratios are calculated from the normalized values. Simple normalization works well in simple cases and less well with complex background problems and complex experimental systems.

### *Z normalization*

As shown above in simple normalization, ratio scores of normalized intensity arrives at a mean to form a ratio to each array. A smaller number of array datasets may lead to large variation. Manipulation of these values between arrays may produce spurious results. Relative intensities between arrays can depend upon factors such as exposure time, labeling efficiency, and hybridization/washing conditions. We have adapted a z-score normalization method based upon a z ratio for a difference between means (Guilford and Fruchter, 73, p. 153) to normalize each array.

In brief, this method involves calculating a distribution of z-scores for all genes in each independent array which sets the mean gene expression as 0 and the standard deviation to 1 by definition. The z normalization method then uses the differences of the z-scores between two or more arrays to calculate differences in gene expression. We have found an empirical relationship between z-ratios and traditional ratios by the following formula:

$$\mathbf{z\text{-ratio} = 0.9 (\log_e \text{traditional fold}).}$$

Using this formula, a traditional 2 fold change equals approximately a 0.8 z-ratio,

5 = ~1.72 z-fold, 10 = ~2.41 z-fold. One can see that z-ratio is more conservative than traditional ratios. Z normalization may allow greater comparability of array results across experiments and array formats.

The formula to calculate a z-ratio follows simple steps that can be applied in a worksheet. A log<sub>10</sub> transformation of the raw intensity data is applied to reduce the variance due to extreme values. The mean and the standard deviation of the log<sub>10</sub> scores for each sample are calculated and entered into a z-score normalization formula:

$$\mathbf{\text{Observed Gene z-score} = (\text{Observed Gene log}_{10} \text{intensity} - \text{Mean microarray all genes log}_{10} \text{intensity}) / (\text{standard deviation microarray all genes log}_{10} \text{intensity}).}$$

Gene expression differences between two arrays are calculated by taking the difference between:

$$\mathbf{z \text{ score difference gene 1} = [(z_{S1a} + z_{S1b})/2] - [(z_{C1a} + z_{C1b})/2]}$$

where S1 and C1 = experimental gene 1 and control gene 1, respectively and a,b, represent individual z-scores obtained from 2 measurements of the gene.

To facilitate comparison to traditional fold differences, the z-score differences are further translated into a z-ratio (Guilford and Fruchter, 1973) based upon the formula:

$$\mathbf{z \text{ score difference gene 1/standard deviation of the z differences distribution}}$$

Depending on the level of false positives that one is willing to accept in these comparisons, a threshold of 0.01, 0.05, 0.1 etc can be adopted. **Replication of the results is found to be of the greatest value in producing believable data rather than relying upon statistical corrections.** When a gene difference replicates on separate arrays, this can certainly lead to further confidence. By chance ~5% of gene differences could arise by chance when adopting p = 0.05 threshold. When probing 1000 genes on an array, one would expect either up- or down- regulation of 50 genes by chance. However, replication of the experiment would allow (0.05)<sup>2</sup> or 2.5 genes to be differentially regulated. If the experiment is repeated a third time, then the probability of finding a significant gene difference by chance drops to only (0.05)<sup>3</sup> or 0.12 genes on a 1000 gene array. Thus, two or three levels of replication can save time and effort before conducting additional validation studies.

### *Correlation coefficient (r), Coefficient of determination (r<sup>2</sup>), and Coefficient of variation (CV)*

Another useful method for determining the reliability of array data, is to either generate scatter plots of duplicate values if available within an array, and between arrays. Once a scatterplot is made, a linear fit of the data by the use of the correlation coefficient (r) can be applied using least squares regression solution which is available in spreadsheet programs and statistical packages.

The square of the correlation coefficient (r<sup>2</sup>) is also known as the coefficient of determination and is useful for estimating how well the data on average is explained by a linear fit between the duplicate array values. The higher the (r<sup>2</sup>) value indicates better reliability of the data around a straight line. The (r<sup>2</sup>) value can range from 0 to 1 with 0 the worst and 1 the best reliability. The (r<sup>2</sup>) value can then be applied to determine whether arrays are giving expected duplications and to examine whether improvements in experimental conditions might give more reproducibility.

An alternative method is to examine the coefficient of variation (CV) within an array for reliability between duplicate values of a gene. The CV is calculated by taking the (standard deviation of duplicates / mean of the duplicate values) and can be multiplied by 100 for expressing the CV as a percentage. Using the CV with 2 array points is useful as a filter for determining whether the data exceeds thresholds set by the user. A CV <20.0% is expected for duplicate values and most often encountered in the range of 2 –F 10% in microarrays. Values of CV that are consistently above this range would lead to concern and investigation of printing, data acquisition, and other issues.

In summary, these three methods are quick ways of exploring the goodness of fit between individual data points and should be routinely monitored for each array experiment.

### *Differential reliability*

Differential reliability is another method that looks at the relative intensity of gene expression and calculates the reliability based upon intervals of gene expression levels. From this standpoint, the smaller variation associated with higher intensity levels, would bring less error in the calculations of the z-ratios and CVs. Conversely, low gene expression levels would have a larger CV associated with the interval.

## **6. Data analysis and visualization**

Once the intensity data is acquired from nylon membranes based array experiments and normalized, further data analysis is very similar to analysis derived from other array formats. While some quantitation programs are stand-alone, other software programs designed for data acquisition of intensity values from radioactive images have normalization, statistical, and display functions as well ( P-Scan, Imagene, etc.). Data organization and visual display challenges are similar for membrane based arrays as other array formats. Clustering, self-organizing maps, multi-dimensional scaling, and hierarchical tree displays can be performed on this type of data as well. WWW links to image processing, data acquisition, statistical, data mining, and visual display software are shown in Appendix 1.

## **7. Conclusion**

cDNA arrays are rapidly having an impact in many areas of biomedical research. Nylon membrane based microarrays offer alternatives that may in some cases be more sensitive, flexible, inexpensive, and universal as compared to other array formats. In some situations of limited RNA or exotic species, membrane arrays may be the most practical experimental approach. As genomic-scale arrays are applied with greater frequency in neuroscience, longstanding questions may be addressed concerning neurodegeneration, psychiatric and developmental disorders, and other complex questions in the brain.

## 8. References

- Barrett, T., Cheadle, C., Wood, W.H., Donovan, D.M., Freed, W.J., Becker, K.G., and Vawter, M.P. Gene Expression Patterns in Neuronal Cells and Tissues Using a Neuronal Specific cDNA Microarray. In preparation.
- Bertucci F, Bernard K, Llorca B, Chang YC, Granjeaud S, Birnbaum D, Nguyen C, Peck K, Jordan BR Sensitivity issues in DNA array-based expression measurements and performance of nylon microarrays for small samples. *Hum Mol Genet* 1999 Sep;8(9):1715-22
- Colantuoni C, Purcell AE, Bouton CM, Pevsner J High throughput analysis of gene expression in the human brain. *J Neurosci Res* 2000 Jan 1;59(1):1-10
- DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14:457-460.
- Duggan DJ, Bittner M, Chen Y, Meltzer P, Trent JM (1999) Expression profiling using cDNA microarrays. *Nat Genet* Jan;21(1 Suppl):10-14.
- Guilford, J.P., Fruchter, B. (1973). *Fundamental statistics in psychology and education*. Fifth edition. McGraw-Hill: New York.
- Khan J, Simon R, Bittner M, Chen Y, Leighton SB, Pohida T, Smith PD, Jiang Y, Gooden GC, Trent JM, Meltzer PS Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res* Nov 15;58(22):5009-5013
- Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG Gene expression profiles of laser-captured adjacent neuronal subtypes. *Nat Med* 1999 Jan;5(1):117-22
- Sargent, T. D. and Dawid, I. B. (1983) Differential gene expression in the gastrula of *Xenopus laevis*. *Science*, 222(4620), 135-9.
- Schena, M., Shalon, D., Davis, R. W. and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, 270(5235), 467-70.
- Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzog Normalization strategies for cDNA microarrays. *Nucleic Acids Res* 2000 May 15;28(10)
- Tanaka, T.S., Jaradet, S. A., Lim, M.K., Kargul, G.J., Wang, X., Grahovac, M.J., Pantano, S., Sano, Y., Piao, Y., Nagaraja, R., Doi, H., Wood, W.H., Becker, K.G., and Ko, M.S.H. Genome-wide expression profiling of mid-gestation placenta and embryo using a 15K mouse developmental cDNA microarray. *Proc Natl Acad Sci U S A* 2000 Aug 1;97(16):9127-32.
- Vawter, M.P., Barrett, T., Cheadle, C., Sokolov, B.P., Wood, W.H., Donovan, D.M., Webster, M., Freed, W.J., Becker, K.G. (2000). Application of cDNA microarrays to examine gene expression differences in schizophrenia. *Brain Research Bulletin* [In Press].
- Whitney LW, Becker KG, Tresser NJ, Caballero-Ramos CI, Munson PJ, Prabhu VV,
- Trent JM, McFarland HF, Biddison WE Analysis of gene expression in multiple sclerosis lesions using cDNA microarrays. *Ann Neurol* 1999 Sep;46(3):425-8
- Zhao, N., Hashida, H., Takahashi, N., Misumi, Y. and Sakaki, Y. (1995) High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. *Gene*, 156(2), 207-13.

## 9. Appendix 1

The following computational and visualization tools are being used in the acquisition and analysis of cDNA array data and complex gene expression patterns in our lab.

### Image Processing/ Data Acquisition

ImageQuant	<a href="http://www.mdyn.com/products/ImageQuant/default.htm">http://www.mdyn.com/products/ImageQuant/default.htm</a>
IPLab	<a href="http://www.scanalytics.com/sos/product/gen/IPLab.html">http://www.scanalytics.com/sos/product/gen/IPLab.html</a>
Adobe Photoshop	<a href="http://www.adobe.com/products/photoshop/main.html">http://www.adobe.com/products/photoshop/main.html</a>
Adobe Illustrator	<a href="http://www.adobe.com/products/illustrator/main.html">http://www.adobe.com/products/illustrator/main.html</a>
Imagene	<a href="http://www.biodiscovery.com/products/ImaGene/imagene.html">http://www.biodiscovery.com/products/ImaGene/imagene.html</a>
DeArray– YiDong Chen NHGRI	<a href="http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html">http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/img_analysis.html</a>
PSCAN– Peter Munson CIT, NIH	<a href="http://absalpha.cit.nih.gov/pscan/">http://absalpha.cit.nih.gov/pscan/</a>

### Data Storage and Manipulation

Microsoft Excel	<a href="http://www.microsoft.com/catalog/display.asp?site=797&amp;subid=22&amp;pg=1">http://www.microsoft.com/catalog/display.asp?site=797&amp;subid=22&amp;pg=1</a>
Microsoft Access	<a href="http://www.microsoft.com/catalog/display.asp?site=769&amp;subid=22&amp;pg=1">http://www.microsoft.com/catalog/display.asp?site=769&amp;subid=22&amp;pg=1</a>
Axum	<a href="http://www.mathsoft.com/axum/">http://www.mathsoft.com/axum/</a>

### Data Visualization and Analysis

Spotfire	<a href="http://www.ivee.com/">http://www.ivee.com/</a>
Cluster	<a href="http://rana.stanford.edu/software/">http://rana.stanford.edu/software/</a>
Tree view	<a href="http://rana.stanford.edu/software/">http://rana.stanford.edu/software/</a>
Gene Spring	<a href="http://www.sigenetics.com/">http://www.sigenetics.com/</a>
MAExplorer–Peter Lemkin NCI/FCRDC, NIH	<a href="http://www-lecb.ncifcrf.gov/mae/maeDoc.html">http://www-lecb.ncifcrf.gov/mae/maeDoc.html</a> <a href="http://www-lecb.ncifcrf.gov/mae/">http://www-lecb.ncifcrf.gov/mae/</a>
MineSet- Silicon Graphics	<a href="http://www.sgi.com/software/mineset/">http://www.sgi.com/software/mineset/</a>