

Differential Gene Expression Patterns Revealed by Oligonucleotide Versus Long cDNA Arrays

Jiang Li,* Matthew Pankratz,† and Jeffrey A. Johnson*†‡§¹

*School of Pharmacy, †Waisman Center, ‡Molecular and Environmental Toxicology Center, and §Center for Neuroscience, University of Wisconsin, Madison, Wisconsin 53705

Received May 29, 2002; accepted July 10, 2002

DNA microarrays can be classified into oligonucleotides (Affymetrix) or long cDNAs (IncyteGenomics) based on the arrayed probes. Unfortunately, data are lacking on the comparison of these two popular global screening array systems. The present study was designed to assess the reliability of datasets generated by the two platforms from the same samples. We have already established a model for upregulation of a cluster of antioxidant responsive element (ARE)-driven genes in a human neuroblastoma cell line by treatment with *tert*-butylhydroquinone (tBHQ) for 8 and 24 h. HuGene FL (Affymetrix), U95 Av2 (Affymetrix), and UniGem V 2.0 (IncyteGenomics) were chosen to do the comparative study on 8- and 24-h samples. The Affymetrix data generated from U95Av2 chips demonstrated that the mRNA of 218 (2.3% of total clones) genes was increased after 8 h of tBHQ treatment. This list included most of the known ARE-driven genes, and nine selected genes showed a high consistency with RT-PCR results. IncyteGenomics called four genes increased and no genes were decreased. These same four genes were also called by the Affymetrix microarray. The sensitivity (fluorescence intensity) and specificity (fold) were very different for selected genes when comparing the two platforms. Cross-hybridization was shown to partially contribute to the discrepancies of the data generated by the two platforms. According to our results, the data generated from oligonucleotide microarrays is more reliable for interrogating changes in gene expression than data from long cDNA microarrays.

Key Words: DNA arrays; gene expression; antioxidant responsive element; cross-hybridization; *tert*-butylhydroquinone.

Recent technological breakthroughs have changed gene expression analysis from the traditional notion of “one gene, one experiment” to a completely parallel, automatic, and miniaturized assay called microarray or genechip. The development of microarrays or genechips has given scientists the ability to assess changes in thousands of genes simultaneously in one succinct set of experiments. More than 2000 papers involving microarray analyses have been published, and most within the last two years (see Fig. 1A). The recent popularity of microar-

ray technology can be attributed to its successful application to a wide range of topics including toxicity profiling, drug screening, and genomic characterization of disease (Bartosiewicz *et al.*, 2000; Harrington *et al.*, 2000; Heller *et al.*, 1997; Marcotte *et al.*, 2001). Despite being a relatively new technique, arrays exist in a variety of forms and can be classified based on any number of attributes, including length of target sequence (long cDNAs or oligonucleotides), commercial or custom made, global or specific, glass- or membrane-based, and spotted or *in situ* synthesized. Because many researchers are interested in the application of microarrays in their systems, many commercial suppliers have made available standard microarrays and analysis packages. At present there are two popular array systems: spotted long cDNA microarrays and *in situ* synthesis oligonucleotide microarrays, which have been developed by several major companies. The platforms differ in array manufacturing and design, array hybridization, scanning, and data handling (Lipshutz *et al.*, 1995; Lockhart *et al.*, 1996; Mirmicz *et al.*, 2000; Okamoto *et al.*, 2000; van Hal *et al.*, 2000). Affymetrix arrays involve *in situ* synthesis of oligonucleotides on glass, whereas other companies such as IncyteGenomics and BD Clontech use long cDNAs derived from ESTs, cloned cDNA, or PCR products attached to a variety of support structures, including nylon filters, glass slides, or silicon chips. Commercial arrays also vary in the number of genes they are able to screen, with some capable of global screening and others only specific functional screening. For example, the Affymetrix human U95 Av2 GeneChip allows for detection of 9670 genes/EST clusters on one array and the Incyte human UniGem V 2.0 allows for detection of 8556 genes/EST clusters on one array. In contrast, Clontech provides small-scale arrays that are particularly designed for toxicology, apoptosis, and cancer research.

Tert-butylhydroquinone (tBHQ), a metabolite of the widely used food antioxidant, butylated hydroxyanisole (BHA), shows significant phase-II enzyme induction, even in very low concentrations (10 nM) in IMR-32 human neuroblastoma cells (Lee *et al.*, 2001a,b). Increasing evidence indicates that tBHQ induces phase-II detoxifying enzymes through the antioxidant responsive element (ARE; Hayes and McMahon, 2001; Tala-

¹ To whom correspondence should be addressed at the School of Pharmacy, University of Wisconsin, 6125 Rennebohm Hall, 777 Highland Ave, Madison, WI 53706. Fax: (608) 262-5345. E-mail: jajohnson@pharmacy.wisc.edu.

lay, 2000). The binding of a transcription factor, NF-E2 related factor 2 (Nrf2), to the ARE leads to activation of a score of genes such as NAD(P)H:quinone oxidoreductase (NQO1), heme oxygenase 1 (HO1), multiple forms of glutathione *S*-transferase (GST), glutathione reductase (GR), and thioredoxin reductase (TR) (Jaiswal, 2000; Itoh *et al.*, 1999; McMahon *et al.*, 2001). Two recent publications from our laboratory have identified the tBHQ-regulated genes in IMR-32 cells using Affymetrix chips (HG U95Av2; Li *et al.*, 2002; Li and Johnson, 2002). Concurrent with these studies, we also had the same sample analyzed using long cDNA arrays (IncyteGenomics). The lack of comparative data using both global screening platforms leaves the researcher to select one platform, with little research directly addressing which platform gives the most accurate dataset. Thus, the present study was designed to identify and compare the tBHQ-induced gene profiles revealed by these two global screening systems.

MATERIALS AND METHODS

Cell culture and mRNA purification. IMR-32 human neuroblastoma cells were grown in DMEM supplemented with fetal calf serum (10%), 100 iU/ml penicillin and 100 mg/ml streptomycin. Cultures were maintained at 37°C in a humidified 10% CO₂ atmosphere. When cells reached 80% confluence, *tert*-butylhydroquinone (tBHQ, Fisher) at a final concentration of 10 mM or vehicle (0.01% EtOH) was added into the medium for 8 or 24 h prior to harvesting total RNA by using Trizol (Gibco BRL, Grand Island, NY). Polyadenylated mRNAs were isolated with Oligotex (Qiagen, Valencia, CA) and were quantified by UV spectrophotometry. All treatment and preparations were performed in triplicate.

Microarray Analysis

cDNA array system. mRNA was sent to IncyteGenomics for analysis on the Incyte Unigem V2.0 arrays (containing 8556 genes/EST cluster). Since Cy3 and Cy5 have a different labeling efficiency for target cDNAs, we reversed the samples for different labeling each time, to determine the reproducibility of the results. Note that the operators were blind to the specific category to which each sample belonged. Following the hybridization and washing, the relative expression level of both cDNA populations were measured and compared by making the balanced Cy3/Cy5 fluorescence ratio. In this study, genes were considered differentially expressed if the ratio was ≥ 1.74 -fold, at 99% confidence (<http://www.incyte.com>). The cDNAs corresponding to genes were sorted by enzyme, function, or pathway cluster analysis using Gemtools software (V2.4.2, IncyteGenomics). Gene group data were exported to MS Excel for further analysis.

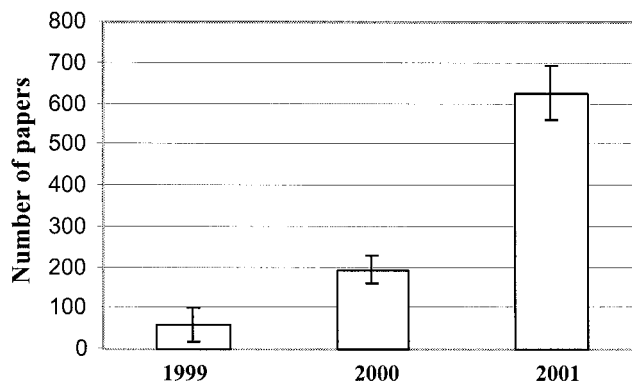
Oligonucleotide array system. The samples analyzed by Incyte were also used to generate fragmented cRNA probes for Affymetrix GeneChip analysis. Briefly, cDNA was synthesized from 500 ng mRNA by using superscript choice kit (GIBCO/BRL) with a T7-(dT)₂₄ primer incorporating a T7 RNA polymerase promoter. The cRNA was prepared and biotin-labeled by *in vitro* transcription (Enzo Biochem). Labeled cRNA was fragmented by incubation at 94°C for 35 min in the presence of 40 mM Tris-acetate, pH 8.1, 100 mM potassium acetate, and 30 mM magnesium acetate. Fifteen mg of the fragmented cRNA was hybridized for 16 h at 45°C to a HuGene FL (original version, 6800 genes/EST cluster) or HG U95Av2 array (updated version, 9670 genes/EST clusters). After hybridization, the genechips were automatically washed and stained with streptavidin-phycoerythrin using a fluidics station. Arrays were finally scanned at 3-mm resolution using the Genechip System confocal scanner made for Affymetrix by Aligent. Microarray Suite 4.1 and 5.0 software from Affymetrix were used on HuGene FL and HG U95Av2 arrays,

respectively, to determine the relative abundance of each gene, based on the average difference of intensities. Output from the genechip analysis was merged with the Unigene or GenBank descriptor and stored as an Excel data spreadsheet.

RT-PCR

We used RT-PCR to validate the increased mRNA level of nine selected genes. PCR primers specific for the genes of interest were used for cDNA synthesis and amplification as follows. Unique oligonucleotide primer pairs for NQO1, HO1, GR, aldo-keto reductase family 1, γ -glutamylcysteine ligase regulatory and catalytic subunits (GCLR and GCLC), thioredoxin reductase (TR), neurofilament heavy subunit, and β -actin were prepared by IDT (Coralville, IA). Total RNA, purified from cell pellets with Trizol Reagent (GIBCO/BRL), was subjected to RT-PCR with Promega Transcription System (Madison, WI). The reaction mix (20 ml) contained 200 mM dNTP, 0.45 mM of each primer, and 1 mg of total RNA and AMV Reverse Transcriptase (15 U). RNA was reverse-transcribed at 42°C for 30 min and DNA was amplified by an initial incubation at 94°C for 4 min, followed by ~ 25 – 35 cycles at 94°C

A



B

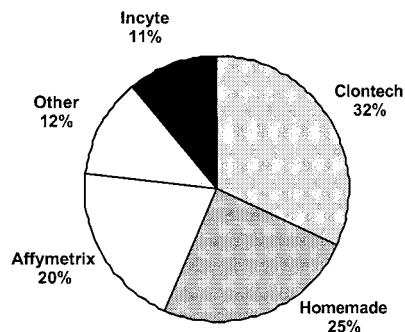


FIG. 1. Statistical analysis of the original articles associated with microarray. A. The number of articles containing microarray data during 1999–2001; B. Types of microarrays used during ~ 1999 – 2001 . Based on 1999, 2000, and 2001, 4-month samples ($n = 272$). Homemade arrays are custom spotted by the investigator and are usually the long cDNA variety. Others are groupings of commercial array companies, including Research Genetics, Sigma GenoSys, Super Array, Rosetta Inpharmatics, Takara, R&D Systems, NEN Life Sciences and Phase-1 Molecular Toxicology.

for 0.5 min, ~55–58°C for 0.6 min, 72°C for 0.5 min, and a final extension at 72°C for 7 min. The PCR products were then separated by electrophoresis in a 1.2% agarose gel and visualized by ethidium bromide staining. The number of cycles and melting temperature were adjusted, depending on the gene being amplified.

RESULTS

Statistical analysis of publications associated with microarray application. A search of the National Institute of Health's PubMed database was performed, using the key words "gene chip," "microarray," or "gene array." Only articles that directly used microarray technology to test a hypothesis were used in the analysis, and these included ones that used arrays to explore molecular phenotyping, functional genomics, pharmacogenomics, toxicogenomics, developmental biology, and sequencing. Other articles that either summarized or analyzed results of existing microarray studies or presented improvements to a technical aspect of performing, creating, or analyzing microarrays were not included. The average and standard deviation of a 4-month sample (January, April, July and October) were calculated and multiplied by a factor of 12 to approximate the average number of articles involving microarray technology for each year. All of these articles also were classified based on the company that produced the arrays used.

Figure 1A illustrates the dramatic increase in the number of publications using microarrays since 1999 (1999, 60 ± 41.6 ; 2000, 195 ± 34.5 ; 2001, 627 ± 66.7) and Figure 1B shows that in the last three years 75% of the arrays used have been provided commercially. Some reasons for the recent increase in microarray use are certainly the improvement in cost, convenience, availability, and performance of commercial array systems. It should also be noted that either Affymetrix or IncyteGenomics provided the vast majority of commercial arrays used for global screening. Only one-fourth of the publications worked on homemade arrays that usually were designed to meet the specific interest of researchers. This statistical analysis demonstrates the necessity of this study to compare these two global commercial array systems, which constitute over 30% of the total arrays used in the last three years and somewhere between 75 and 90% of the total global screening arrays used.

Distinct Gene Expression Patterns Revealed by the Two Array Systems

Oligonucleotide array system. The Affymetrix algorithm for detecting differential expression considers several parameters from the raw data, and the results of these analyses are presented in the Supplemental Appendix of Affymetrix Microarray Suite User Guide. "Difference Call" or "Change Call" was used to determine the differential gene expression initially. Original (HuGene FL) and updated versions (HG U95Av2) of human chips were analyzed by Microarray Suite V. 4.1 and V. 5.0, respectively. The data generated from HuGene FL chips

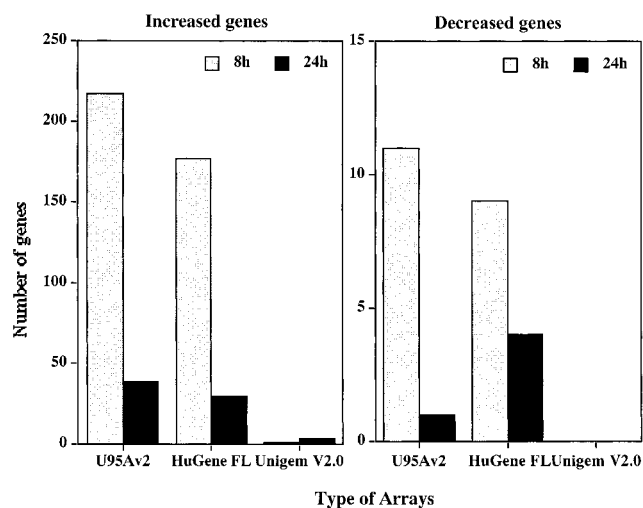


FIG. 2. Ranking analysis demonstrated the numbers of genes, whose mRNA levels were increased or decreased after treatment with tBHQ for 8 and 24 h, from the microarray data generated by the two platforms. HuGene FL (Affymetrix), HG U95Av2 (Affymetrix), and Unigem V. 2.0 (IncyteGenomics) chips were used to do the comparative study.

has a high correlation with that from the U95Av2 chips. There were dramatic variations in the number of genes called different (I, increased; D, decreased) calculated from the matched sample pairs. For example, the number of I/D genes generated from three pair-matched, 8-h treatment samples on the U95Av2 chips were 173/283, 171/151, and 597/300. We defined increased, decreased, or no change of expression for individual genes based on ranking of the "difference call" or "change call" from the three pair-matched comparisons. Briefly, No change, 0; Marginal increase, 1; Increase, 2; Marginal decrease, 1; and Decrease, 2. The cutoff value for increase/decrease was set as ± 3 because of the marginal calls. This screening process led to the identification of a certain number of genes whose expression levels consistently increased or decreased. Application of these principles to the other sample pairs led to the identification of approximately 218 (2.3% of total) genes on the U95Av2 chip whose expression consistently increased (207) or decreased (11) after tBHQ treatment for 8 h. The number of increased or decreased genes on the HuGene FL and HG U95Av2 chips were calculated and are shown in Figure 2.

One of the major clusters of increased genes included many of the Phase-II detoxification enzymes, such as early-response genes (NQO1, HO1, GR, GSTM3, GCLR, GCLC, and TR), and late-response genes (transketolase, ferritin heavy and light chains; Table 1). Other antioxidant systems like hepatic dihydrodiol dehydrogenase (aldo-keto reductase family 1) and its isoform, KIAA0119 (a subtype of aldo-keto reductase family 1, AKR1C3), malate NADP oxidoreductase, and breast cancer cytosolic NADP(+)-dependent malic enzyme were also found to be increased by tBHQ. Some have been verified to be ARE-driven genes whereas others have been proposed to be

TABLE 1
Summary Statistics of Gene Expression Data Generated by Two Microarray Systems

Gene name	Incyte UniGem V 2.0			Affymetrix HuGene FL				Affymetrix U95Av2					
	Accession no.	8 h (n = 5)	24 h (n = 1)	Accession no.	8 h (n = 3)		24 h (n = 3)		Accession no.	8 h (n = 3)		24 h (n = 3)	
		Fold	Fold		Fold	Call	Fold	Call		Fold	Call	Fold	Call
Thioredoxin reductase	D88687	2.04	2.70	X91247	1.95	I	2.47	I	X91247	2.56	I	2.20	I
NAD(P)H: quinone oxidoreductase	M81600	1.04	1.20	J03934	10.6	I	14.8	I	M81600	5.40	I	5.45	I
Neurofilament heavy	X15306	1.26	1.20	X15306	2.5	I	1.97	I	X15306	2.27	I	1.52	NC
Transketolase	AW006207	1.30	1.20	L12711	1.2	NC	2.20	I	L12711	0.85	NC	1.63	I
Breast cancer cytosolic NADP(+)-dependent malic enzyme	N/A	N/A	N/A	U43944	2.10	I	5.10	I	U43944	3.26	I	2.16	I
Malic enzyme 1 NADP(+)-dependent, cytosolic	NM_00239	1.72	1.60	N/A	N/A	N/A	N/A	N/A	AL049699	11.06	I	5.98	I
Aldo-keto reductase family 1, member C4 (AKR1C4)	S68287	1.60	4.60	U05861	4.85	I	5.13	I	U05861	3.57	I	17.78	I
Aldo-keto reductase family 1, member C1 (AKR1C1)	M86609	1.10	1.80	D17793	6.35	I	4.13	I	D17793 (AKR1C3)	13.0	I	16.2	I
Ferritin, light chain	BE301211	1.14	2.30	M11147	1.35	NC	2.33	I	AL031670	-0.41	NC	1.66	I
Ferritin, heavy chain	A1816415	1.46	1.70	L20941	1.35	NC	2.57	I	L20941	0.97	NC	2.15	I
Heme oxygenase 1	Z82244	1.14	1.3	Z82244	4.55	I	1.70	NC	Z82244	7.38	I	2.74	NC
γ -Glutamylcysteine ligase, regulatory subunit	NM_00206	1.04	1.30	L35546	7.50	I	4.47	I	L35546	4.99	I	3.74	I
γ -Glutamylcysteine ligase, catalytic subunit	AL033397	1.20	1.26	M90656	1.40	NC	1.38	NC	M90656	1.33	I	0.94	I
Glutathione reductase	N/A	N/A	N/A	X15722	2.2	MI	3.73	I	X15722	4.96	I	1.97	NC

Note. Data analysis was performed across all pair-matched tBHQ and vehicle-treated groups. The fold call and difference call (or change call) in Affymetrix HuGene FL and U95 Av2 chips were calculated by Microarray Suite 4.1 and 5.0, respectively, and the fold changes were presented as mean. Each transcript in the comparison analysis had 5 possible difference call (or change call) outcomes: (1) increase, I; (2) marginally increase (MI); (3) decrease (D); (4) marginally decrease (MD); and (5) no change (NC). The folds in Incyte Unigem V.2.0 arrays were calculated by the Incyte algorithms and the results were presented as mean. N/A, representing the gene probe, was not available in that array.

upregulated through ARE activation (Hayes and McMahon, 2001; Itoh *et al.*, 1999; Jaiswal, 2000; Li and Johnson, 2002; Li *et al.*, 2002; McMahon *et al.*, 2001; Talalay, 2000). Differential expression of eight selected genes was verified by RT-PCR (Fig. 3).

cDNA array system. Since the updated HG U95Av2 gene-chip covered a similar number of genes, many of which also were covered on the Unigem V2.0 array, the data generated from the U95Av2 chips was compared to that from the Unigem V2.0 array. The data generated from 5 repeats on 8-h sample pairs were analyzed (3 sets Vehicle-cy3, Treatment-cy5; 2 sets Vehicle-cy5, Treatment-cy3). Of all genes and expressed sequence tags (ESTs) probed by the five microarrays, an average of 3.0 ± 2.0 (increased) and 1.6 ± 1.0 (decreased) gene transcripts per array were judged to be differentially expressed, based on the criterion of having a balanced ratio ≥ 1.74 -fold (99% confidence level). An "increased" or "decreased" call was assigned to these genes. A reproducible dataset was generated based on ranking the calls and summing up the ranks of each gene. The final gene list was determined by sorting the total rank over 2 in 5 repeats. Of the 8556 genes analyzed by Incyte, only one gene, TR with a 2.04-fold increase, satisfied

the inclusion criteria at 8 h of treatment (Table 1). A 24-h sample pair was also analyzed; however, only three other genes were added to the list (aldo-keto reductase family 1, member C1 and C4, and ferritin light chain; Table 1). Ferritin heavy chain (1.70-fold) was suspected since its balanced ratio ranged around the threshold. The number of genes finally picked up by the Incyte arrays is shown in Figure 2.

Cross-hybridization. Sensitivity and specificity are the two major features that researchers are concerned with in performing microarray analysis. Microarray sensitivity is often defined as the minimum reproducible signal (fluorescence intensity) detected by a given array-scanning system. The sensitivity (fluorescence intensity) and specificity (fold) was very different in the same genes when comparing the two platforms. For example, HO1 and NQO1 changed 7.38-fold and 5.40-fold on the U95Av2 chips, respectively, after 8 h of treatment, whereas the cDNA array showed 1.14-fold and 1.04-fold change for both genes (Table 1). Cross-hybridization must be considered since false positive signals sometimes disguise real changes and make the basal expression of individual genes very high. This can lead to a false impression of high sensitivity, albeit this high sensitivity is reproducible. For example, labeling and

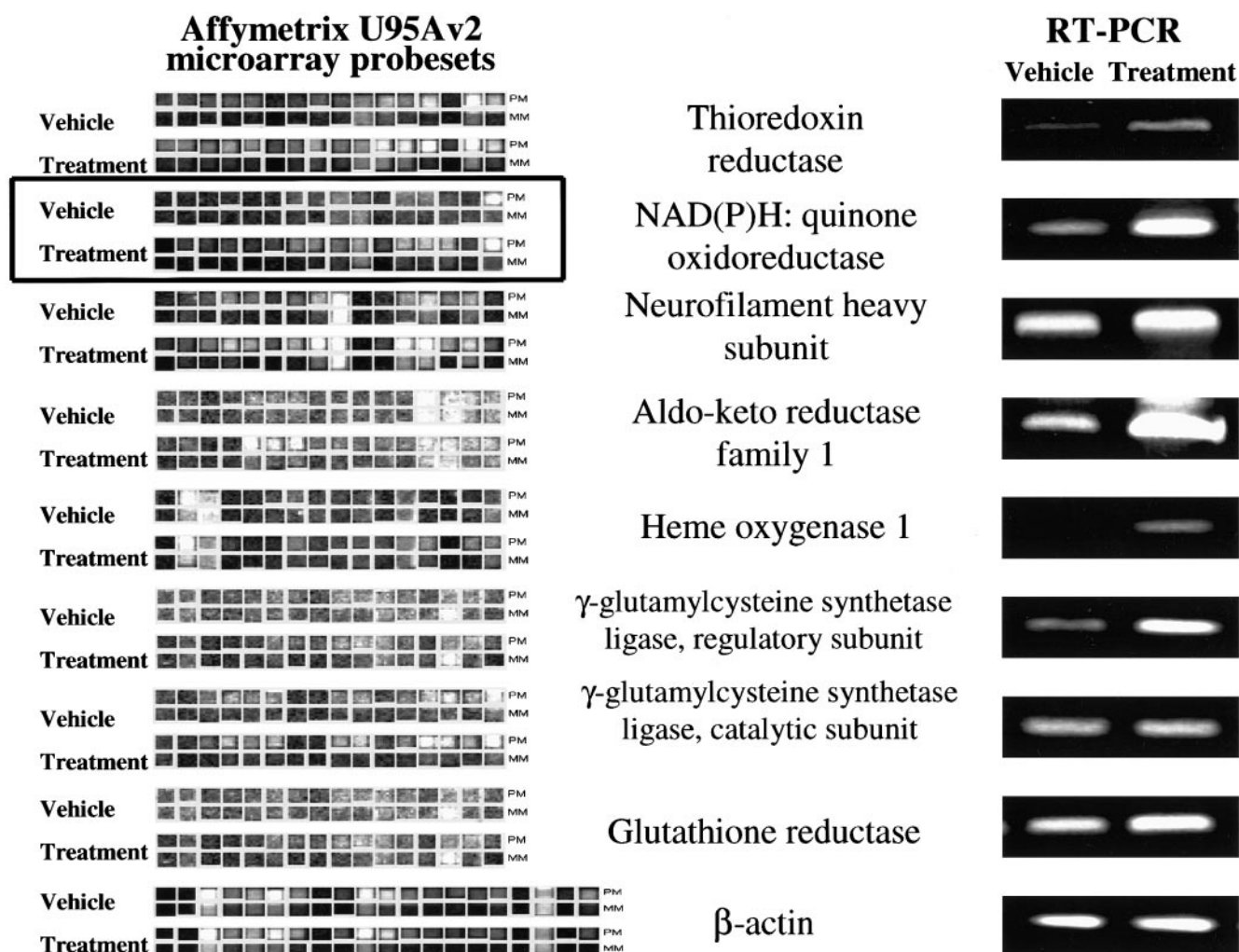


FIG. 3. Differentially expressed genes in tBHQ and Vehicle treated IMR-32 cells detected by Affymetrix-based microarray analysis and verified by RT-PCR. Individual arrays were hybridized with biotinylated cRNA probes generated from IMR-32 cells treated with tBHQ or vehicle (0.01% EtOH) for 8 h. The image showed eight of the genes, thioredoxin reductase, NAD(P)H: menadione oxidoreductase (NQO1), neurofilament heavy subunit, aldo-keto reductase family 1, heme oxygenase 1, γ -glutamylcysteine synthetase regulatory and catalytic subunits, and glutathione reductase, which were identified as being transcriptional targets of the Nrf2. The grid images for β -actin were used as control.

hybridization of IMR-32 cell cRNA with U95Av2 chips did not result in a detectable signal for some probe sets designed to detect certain phase-II detoxification enzymes such as NQO1 and HO1. Both were called absent in controls (Table 2). After treatment with tBHQ, there was a significant increase in signal (fluorescence intensity) of these genes and both were called present (Table 2). The Incyte 2-color analytic strategy provided the relative probe value of gene expression, which seems to be aberrantly high for NQO1 and HO1, both in control and treated groups. Thus, the obvious change in NQO1 based on RT-PCR (Li and Johnson, 2002), Western blot (Li and Johnson, 2002), and enzymatic activity (Moehlenkamp and Johnson, 1999) has been buried in the cross-hybridization signal on the spotted arrays. This holds true for genes whose basal gene expression is low (e.g., HO1) or high (e.g., GCLR; Table 2). As a result,

the disparity in the final datasets probably resulted from some cross-hybridization to the longer cDNA probes on the Unigem V.2 arrays.

Similarly, Affymetrix grid images also showed that different probe sets for the same gene could give quite different results (Fig. 3). For example of NQO1, as shown in Figure 4, each probe pair consists of a perfect match oligonucleotide (PM), designed to be perfectly complementary to a reference sequence, and mismatch oligonucleotide (MM), identical in sequence to the PM probe except for a single base difference. The MM probe help account for cross-hybridization in the data analysis. Through PM and MM comparisons, cross-hybridization could significantly contribute to those false positive signals seen on the long cDNA arrays. Fortunately, Affymetrix algorithms take the cross-hybridization into account, and sig-

TABLE 2
Cross-Hybridization Partially Contributes to the Discrepancies of the Data Generated by the Two Microarray Platforms

Gene name	Incyte UniGem V. 2.0 (<i>n</i> = 5)		Affymetrix U95Av2 (<i>n</i> = 3)			
	Vehicle Probe value ^a	Treatment Probe value	Vehicle Signal value ^b	Detection call ^b	Treatment Signal value	Detection call
Thioredoxin reductase	2977 ± 304	6392 ± 621 ^c	15250 ± 3182	P	37787 ± 426	P
Heme oxygenase 1	5851 ± 1052	6733 ± 1662	255 ± 207	A	1834 ± 549 ^c	P
NAD(P)H: quinone oxidoreductase	4976 ± 772	4785 ± 1013	1168 ± 486	A	8261 ± 884 ^c	P
γ-Glutamylcysteine synthetase, regulatory subunit	17380 ± 1371	17410 ± 1056	849 ± 402	P	4713 ± 1322 ^c	P
γ-Glutamylcysteine synthetase, catalytic subunit	17380 ± 1371	17410 ± 1056	5140 ± 328	P	6922 ± 782 ^c	P

^aThe probe value was calculated by the Incyte algorithms and the results were downloaded through Gemtool software.

^bThe signal value and detection call (P, present; A, absent) were calculated by the Affymetrix Microarray Suite software according to algorithms developed and indicated by the manufacturer. Statistical analysis was conducted by one-tail, paired *t*-test and the data were presented as mean ± SD.

^cRepresented a significant difference of signal value between vehicle and treatment groups (*p* < 0.05).

nals generated by cross-hybridization are largely eliminated to calculate the differential gene expression.

DISCUSSION

The use of DNA microarrays for the analysis of biological samples is becoming a mainstream tool in exploring the transcriptional regulation of gene expression. A key step in the regulatory networks that control ARE-driven gene expression is the sequence-specific binding of a transcription factor, Nrf2, to its DNA recognition site(s). A more complete understanding of these DNA-protein interactions will permit a more comprehensive and quantitative mapping of the regulatory pathways, as well as a deeper understanding of the potential functions of new genes revealed by microarray analysis. Other studies have uncovered many known as well as potentially new ARE-driven genes using oligonucleotide microarray (Li *et al.*, 2002; Li and Johnson, 2002). This paper, however, suggests some fundamental disparities in the results generated from the two global screening platforms. The long cDNA array identified only four genes to be transcriptionally upregulated, whereas the oligonucleotide array data provided a much more comprehensive set of differentially expressed genes. Both platforms take great care to assure the efficiency and accuracy of every step in the

microarray analysis, including mRNA extraction, robust probe generation, hybridization, and data mining (Harrington *et al.*, 2000). The arrayed probes in the two systems, however, are quite different in manufacturing and design, array hybridization, scanning, and data handling. These differences could account for the discrepancies observed in the final results (Schulze and Downward, 2001; Watson and Akil, 1999). This must be taken into consideration when deciding between commercial array products for use in an initial global screening because of their impacts on the array specificity and sensitivity.

The key factor in determining the specificity and sensitivity of the two commercial array systems we used was the length of the DNA employed as the probe sequence on the array. For the Incyte long cDNA array, every clone was selected from the UniGem database, which contains sequences that have been verified as representative of the desired gene. The sequence of probes varied from 500–5000 base pairs, averaging 1000 bp (<http://www.incyte.com>) One of the problems associated with having such a large target sequence is that it becomes difficult to control the hybridization efficiencies of various cDNA probes. Since the hybridization conditions are based on the length of nucleic acid fragments and the compositions of G + C and A + T (or A + U), hybridization efficiency can vary

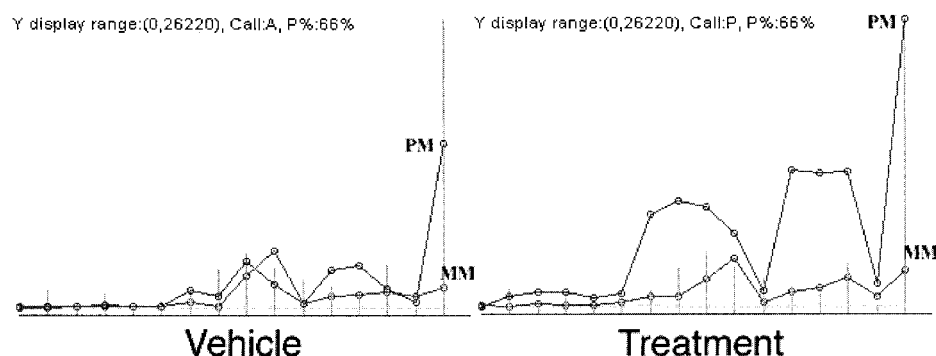


FIG. 4. NQO1 probe levels of PM/MM were illustrated by Dchip software (<http://www.biostat.harvard.edu/complab/dchip>). The graph showed the difference between vehicle (left) and treatment (right) RNA samples for each of the PM and MM probes in the target cluster for NQO1. The *x*-axis ordered probe sets from 1 to 16, and the *y*-axis represented the probe intensities with range from 0 to 26220.

widely when long cDNA sequences are used as a probe. In addition, the issue of cross-hybridization of related or overlapping genes is also potentially an important limitation of the cDNA-based microarray, since different genes may encode common domains and can have some degree of sequence identity or homology with proteins from other genes (Evertsz *et al.*, 2001; Kane *et al.*, 2000). As a result, gene families present a potential problem, because in many cases these genes have a great degree of sequence identity and can only be distinguished from each other by the design and use of gene-specific hybridization probes. In the case of Affymetrix's oligonucleotide array, the 25-oligomer probes are designed to uniquely represent the desired cognate gene through blasting the GenBank database, thus minimizing cross-hybridization between similar sequences (Lipshutz *et al.*, 1995; Lockhart *et al.*, 1996). Using oligonucleotides also makes it easier to design and select probes from the same or different genes with similar G + C content and putative melting temperature. Other criteria for the selection of probes include unique sequence, minimum secondary structures, and 3' terminal sequence selection, all of which serve to insure greater specificity in the probe-target binding (<http://www.affymetrix.com>).

The oligonucleotide approach for the analysis of gene expression has been criticized for a lack of sensitivity (Schulze and Downward, 2001). Since a single capture probe is not always sufficient to distinguish the expression of a particular gene, the use of multiple capture probes to represent a single gene in the Affymetrix array system is intended to avoid the problem of cross-hybridization as well as to increase sensitivity. Thus, one of the likely causes of increased specificity in oligonucleotide arrays, as compared with long cDNA arrays, is a decrease in cross-hybridization with highly homologous genes.

Although it is true that shorter oligonucleotides (15 to 20 mers) promise more specific binding with probes, some papers recently indicated that 50 or 70 mers yield good sensitivity while maintaining the excellent specificity of shorter sequences. Operon Technologies (<http://www.westburg.nl>) compared the sensitivity of 35, 50, 70, and 90-mers for detecting highly expressed genes and genes expressed at moderate or low levels, and found that the 70-mer lengths performed best. However, the results from MWG Biotech (<http://www.mwgbio.com>) showed that 50-mer oligonucleotide arrays provided high specificity and excellent sensitivity, while 70 mers only increased formation of secondary structure (Kane *et al.*, 2000).

Microarray analysis has gone from an interesting idea to a core technology in just a few years. The recent trend of institutions bringing these tools to core facilities should only further open the way for academic researchers to use this innovation. While only a few of the academic core facilities surveyed offered access to the Affymetrix GeneChips, a much larger number offered cDNA array technology. Also based on the number of publications associated with microarray tech-

nology, nearly 75% of papers use spotted cDNA arrays as a means of profiling the gene expression patterns since 1999. In addition, the wide use of the Affymetrix microarray system has been limited, due to the more involved sample processing steps. The investigator is generally responsible for isolating RNA, creating double-stranded cDNA incorporating a T7 promoter, performing IVT to get labeled cRNA, fragmenting the cRNA, and providing QC documentation of these steps. There are also additional costs in that Affymetrix recommends running a test chip to verify the quality of samples.

In the context of a drug-treated cell line presented herein, oligonucleotide arrays give a more accurate and comprehensive scenario of gene-expression profiles, which in turn ensures with a higher probability that further research work will proceed down the correct path. Clearly the recent efforts to genetically profile toxic compounds and promising pharmaceuticals requires the best data sets possible for application to drug discovery. These data strongly imply that oligonucleotide-based arrays are more reliable for global screening compared to long cDNA array at the present time. What the future holds, however, remains to be determined.

ACKNOWLEDGMENTS

We thank Matthew Slattery in the Molecular Biology Facility Core, School of Pharmacy, University of Wisconsin for conducting the Affymetrix gene array hybridization. We thank Thor Stein for reading the manuscript and his very helpful suggestion. This study was supported by grants ES08089 (JAJ), ES10042 (JAJ) and ES09090 (Environmental Health Sciences Center Grant) from the National Institute of Environmental Health Sciences and by the Burroughs Wellcome New Investigator in Toxicological Sciences Award (J.A.J.).

REFERENCES

- Bartosiewicz, M., Trounstein, M., Barker, D., Johnston, R., and Buckpitt, A. (2000). Development of a toxicological gene array and quantitative assessment of this technology. *Arch. Biochem. Biophys.* **376**, 66–73.
- Evertsz, E. M., Au-Young, J., Ruvolo, M. V., Lim, A. C., and Reynolds, M. A. (2001). Hybridization cross-reactivity within homologous gene families on glass cDNA microarrays. *Biotechniques* **31**, 1182, 1184, 1186 passim.
- Harrington, C. A., Rosenow, C., and Retief, J. (2000). Monitoring gene expression using DNA microarrays. *Curr. Opin. Microbiol.* **3**, 285–291.
- Hayes, J. D., and McMahon, M. (2001). Molecular basis for the contribution of the antioxidant-responsive element to cancer chemoprevention. *Cancer Lett.* **174**, 103–113.
- Heller, R. A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D. E., and Davis, R. W. (1997). Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 2150–2155.
- Itoh, K., Ishii, T., Wakabayashi, N., and Yamamoto, M. (1999). Regulatory mechanisms of cellular response to oxidative stress. *Free Radic. Res.* **31**, 319–324.
- Jaiswal, A. K. (2000). Regulation of genes encoding NAD(P)H:quinone oxidoreductases. *Free Radic. Biol. Med.* **29**, 254–262.
- Kane, M. D., Jatke, T. A., Stumpf, C. R., Lu, J., Thomas, J. D., and Madore, S. J. (2000). Assessment of the sensitivity and specificity of oligonucleotide (50-mer) microarrays. *Nucleic Acids Res.* **28**, 4552–4557.

- Lee, J. M., Hanson, J. M., Chu, W. A., and Johnson, J. A. (2001a). Phosphatidylinositol 3-kinase, not extracellular signal-regulated kinase, regulates activation of the antioxidant-responsive element in IMR-32 human neuroblastoma cells. *J. Biol. Chem.* **276**, 20011–20016.
- Lee, J. M., Moehlenkamp, J. D., Hanson, J. M., and Johnson, J. A. (2001b). Nrf2-dependent activation of the antioxidant responsive element by *tert*-butylhydroquinone is independent of oxidative stress in IMR-32 human neuroblastoma cells. *Biochem. Biophys. Res. Commun.* **280**, 286–292.
- Li, J., and Johnson, J. A. (2002). Time-dependent changes in ARE-driven gene expression by use of a noise-filtering process for microarray data. *Physiol. Genomics* **9**, 137–144.
- Li, J., Lee, J. M., and Johnson, J. A. (2002). Microarray analysis reveals an antioxidant-responsive element-driven gene set involved in conferring protection from an oxidative stress-induced apoptosis in IMR-32 cells. *J. Biol. Chem.* **277**, 388–394.
- Lipshutz, R. J., Morris, D., Chee, M., Hubbell, E., Kozal, M. J., Shah, N., Shen, N., Yang, R., and Fodor, S. P. (1995). Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19**, 442–447.
- Lockhart, D. J., Dong, H. L., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680.
- Marcotte, E. R., Srivastava, L. K., and Quirion, R. (2001). DNA microarrays in neuropsychopharmacology. *Trends Pharmacol. Sci.* **22**, 426–436.
- McMahon, M., Itoh, K., Yamamoto, M., Chanas, S. A., Henderson, C. J., McLellan, L. I., Wolf, C. R., Cavin, C., and Hayes, J. D. (2001). The Cap'n'Collar basic leucine zipper transcription factor Nrf2 (NF-E2 p45-related factor 2) controls both constitutive and inducible expression of intestinal detoxification and glutathione biosynthetic enzymes. *Cancer Res.* **61**, 3299–3307.
- Mirics, K., Middleton, F. A., Marquez, A., Lewis, D. A., and Levitt, P. (2000). Molecular characterization of schizophrenia viewed by microarray analysis of gene expression in prefrontal cortex. *Neuron* **28**, 53–67.
- Moehlenkamp, J. D., and Johnson, J. A. (1999) Activation of antioxidant/electrophile-responsive elements in IMR-32 human neuroblastoma cells. *Arch. Biochem. Biophys.* **363**, 98–106.
- Okamoto, T., Suzuki, T., and Yamamoto, N. (2000). Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat. Biotechnol.* **18**, 438–441.
- Schulze, A., and Downward, J. (2001). Navigating gene expression using microarrays—a technology review. *Nat. Cell Biol.* **3**, E190–195.
- Talalay, P. (2000). Chemoprotection against cancer by induction of phase-II enzymes. *Biofactors* **12**, 5–11.
- van Hal, N. L., Vorst, O., van Houwelingen, A. W., Kok, E. J., Peijunburg, A., Aharoni, A., van Tunen, A. J., and Keijer, J. (2000). The application of DNA microarrays in gene expression analysis. *J. Biotechnol.* **78**, 271–280.
- Watson, S. J., and Akil, H. (1999). Gene chips and arrays revealed: A primer on their power and their uses. *Biol. Psychiatry* **45**, 533–543.