

Evaluation of techniques using amplified nucleic acid probes for gene expression profiling

Mehrnoosh Saghizadeh^{a,b}, Donald J. Brown^d, Jian Tajbakhsh^c, Zugen Chen^c, M. Cristina Kenney^d, Debora B. Farber^{a,b}, Stanley F. Nelson^{b,c,*}

^a Jules Stein Eye Institute, UCLA School of Medicine, Los Angeles, CA 90095, USA

^b Molecular Biology Institute, UCLA School of Medicine, Los Angeles, CA 90095, USA

^c Department of Human Genetics, Room 5506b Gonda Center, UCLA School of Medicine, UCLA Medical Center, Los Angeles, CA 90095-7088, USA

^d Department of Ophthalmology, University of California Irvine Medical Center, Orange, CA 92697, USA

Received 31 August 2002; accepted in revised form 7 January 2003

Abstract

Gene expression analyses using spotted cDNA microarrays typically require relatively large quantities of total RNA (up to 100 µg) or polyA+RNA (1–5 µg). However, samples obtained by microdissection, patient biopsies, or embryonic samples often are small and yield an insufficient amount of RNA. Methods such as linear RNA amplification by in vitro transcription (IVT) or cDNA amplification by PCR are currently being used to circumvent these limitations. In the present study, labeled probes from mouse liver and kidney were generated with two amplification methods and were analyzed in terms of reproducibility of intensity values from repeated experiments. In addition, the reliability of differential gene expression detection among the different types of amplified and non-amplified probes was assessed. Data derived from IVT-amplified RNA, as well as from PCR-amplified cDNA probes were reproducible with correlation coefficients of 0.89 and 0.91, respectively. 88–92% of the strongly differentially expressed genes detected with non-amplified probes were also detected as being at least two-folds differentially expressed with the amplified probes. Both the PCR-amplified probe and the IVT-amplified probe were comparable in reproducibility and reliability.

© 2003 Elsevier Science B.V. All rights reserved.

Keywords: DNA microarrays; Gene expression profiling; SMART™ cDNA amplification; T7 amplification; In vitro transcription; Fluorescence labeling; Hybridization

1. Introduction

DNA microarrays are currently used in a variety of biomedical disciplines to explore gene expression on a genome-wide scale. A typical expression profiling experiment involves the hybridization of a fluorescently labeled cDNA probe to cDNA clones immobilized onto glass slides. Usually, 10–100 µg of total RNA or 1–5 µg of polyA+ RNA are used as template to generate sufficient labeled first strand cDNA to perform a single comparative hybridization on a glass microarray containing over 1000 cDNA elements per cm² [1]. A limitation to the application of this technology is the

relatively large amount of RNA required per hybridization. Thus, methods of probe amplification that maintain transcript representation and proportion are needed to circumvent the difficulty in obtaining enough RNA from small samples or rare tissues. Strategies that have been adopted for DNA microarray analyses utilize amplification of full-length double-stranded cDNA via PCR using the SMART system (Clontech, Palo Alto, CA) [2–7], and linear amplification of copy RNA (cRNA) by in vitro transcription (IVT) [8–16]. The SMART system (Switching Mechanism at the 5' end of RNA template) relies on the terminal transferase activity of reverse transcriptase, which adds additional nucleotides to the 3' end of the first strand cDNA when extension has reached the 5' end of the messenger RNA. An additional oligonucleotide included in the reaction hybridizes with the terminal tail of the cDNA, and

* Corresponding author. Tel.: +1-310-794-7981; fax: +1-310-794-5446.

E-mail address: snelson@ucla.edu (S.F. Nelson).

serves as template for the transcriptase to incorporate the primer sequence into the cDNA [17]. Alternatively, signal enhancement has been applied post hybridization to increase the detection sensitivity. This approach includes techniques based on tyramide signal amplification [18–20] or detection using dendrimer technology [21]. Still, probe limitation is a problem with these signal enhancement methods, especially if replicate hybridization experiments need to be performed for statistical validation.

The present study compares the results of hybridizing probes obtained using SMART and IVT to probes generated from non-amplified polyA⁺ and total RNA. These four different techniques for probe production were assessed for reproducibility by repeated experiments using each probe. The average signals from the repeat experiments with each probe were used to detect differentially expressed genes with each method to assess the reliability of differential gene expression.

2. Materials and methods

2.1. Tissue and RNA isolation

Total RNA was isolated from fresh frozen C57Bl/6 mouse liver and kidney using Trizol™ (Invitrogen, Carlsbad, CA), and polyA⁺ RNA was obtained using the Oligotex™ mRNA purification kit (Qiagen, Valencia, CA). The yield and the purity of RNA were determined by calculating the ratio of optical densities measured at wavelengths of 260 and 280 nm.

2.2. SMART-cDNA amplification

Total RNA from each tissue was reverse-transcribed using the SMART PCR cDNA synthesis kit (Clontech Laboratories) as previously described [3]. 1 µg total RNA was mixed with a modified oligo (dT) (CDS primer, final concentration 2 µM), and SMART II oligonucleotide (final concentration = 1 µM), in a volume of 5 µl. The mixture was incubated at 70 °C for 2 min and then immediately chilled on ice. Five microliters of a master mixture composed of 50 mM Tris–HCl (pH 8.3), 6 mM MgCl₂, 75 mM KCl, 1 mM dNTPs, 2 mM dithiothreitol (DTT) and 200 U Superscript II reverse transcriptase (Invitrogen) were added. The reaction was incubated for 1 h at 42 °C, followed by inactivation of reverse transcriptase at 72 °C for 10 min.

The cDNA was diluted 1:5 with 10 mM Tris–HCl (pH 7.6), 1 mM EDTA, and stored at –20 °C until use. One microliter of diluted cDNA was added to 99 µl of 40 mM Tricine–KOH (pH 9.2), 15 mM KOAC, 3.5 mM Mg (OAc)₂, 75 µg/ml bovine serum albumin, 0.2 mM dNTP, 0.2 µM PCR primer (AAGCAGTGGTATCAACGCAGAGT), and 1 × Advantage 2 polymerase

mix (Clontech Laboratories). The cDNA was amplified using the following program profile: one cycle at 95 °C for 1 min; 15 cycles at 95 °C for 20 s and 68 °C for 6 min; followed by one cycle at 72 °C for 10 min. The cDNA was purified with ChromaSpin™ 1000 columns (Clontech) to remove unincorporated primers.

2.3. IVT cRNA amplification

cRNA was prepared from 1 µg total RNA in two steps [8]. In the first step, a T7-bacteriophage promoter was incorporated at the 3' RNA end during the first strand cDNA synthesis. Total RNA was denatured together with 1 µl (100 pmol/µl) oligo-(dT)24-T7 (5'-AAA CGA CGG CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG CGC-3') primer (Genset, La Jolla, CA) at 70 °C for 10 min and put on ice while adding 4 µl of 5 × first strand reaction buffer, 2 µl of 0.1 M DTT, 1 µl of 10 mM dNTP-mix, 300 U Superscript II reverse transcriptase (Invitrogen) and 1 µl (24.800 U/ml) RNAGuard (Amersham Biosciences, Piscataway, NJ) in a total volume of 20 µl. After incubating for 1 h at 42 °C, another 300 U of Superscript II was added, and the reaction mix was incubated for a second hour at 42 °C. Double stranded cDNA was synthesized for 2 h at 16 °C using 30 µl of 5 × second strand reaction buffer, 3 µl of 10 mM dNTP-mix, 10 U of DNA ligase, 40 U of DNA polymerase I, and 2 U of RNase H (Invitrogen) in a total volume of 150 µl followed by addition of 20 U of T4 DNA polymerase (Invitrogen) and a 5 min incubation at 16 °C. The reaction was stopped with 10 µl of 0.5 M EDTA. Double-stranded cDNA was purified with phenol:chloroform:isoamyl alcohol (25:24:1) (Ambion, Austin, TX) extraction using Phase Lock Gels™ light (Eppendorf AG, Hamburg, Germany), precipitated with 0.5 V/V NH₄Ac plus 2.5 V/V 100% ethanol and washed once with 500 µl of 80% ethanol. The cDNA pellet was resuspended in 7 µl of nuclease free water (Ambion) and used for the second step, in which cRNA was synthesized in a transcription assay applying the T7 Megascript kit (Ambion) in a total volume of 20 µl at 37 °C for 4 h. Template DNA was removed by a 15 min incubation at 37 °C with 2 U of DNase I (Ambion). cRNA was purified using the RNeasy kit (Qiagen). All reactions were performed in a DNA Engine Tetrad™ thermocycler with heated lids (MJ Research, Watertown, MA).

2.4. Probe labeling and hybridization

2.4.1. Total RNA and polyA⁺ RNA probes

To generate labeled probes, total RNA or polyA⁺ RNA was reverse-transcribed and labeled directly with Cy3-dUTP or Cy5-dUTP (Amersham Biosciences) using poly-dT as primer (directly labeled cDNA).

Briefly, 70 µg total RNA or 2 µg polyA+ RNA was mixed with 0.5 µg/µl oligo (dT)_{12–18} primer (Invitrogen), in a volume of 12 µl. The mixture was incubated at 70 °C for 10 min, and then chilled on ice. Eighteen microliters of 50 mM Tris–HCL (pH 8.3), 6 mM MgCl₂, 0.5 mM dA/dC/dGTP, 0.1 mM dTTP, 10 mM DTT, 30 U RNasin, 0.1 mM Cy3- or Cy5-dUTP (Amersham Biosciences), and 400 U Superscript II reverse transcriptase were added and the mixture incubated for 1 h at 42 °C. Another 400 U Superscript II reverse transcriptase were added, and the reaction was incubated for another hour at 42 °C. The RNA template was removed by the addition of 0.2 µl RNase A (10 µg/µl) and incubation at 37 °C for 10 min, followed by the addition of 1.5 µl 10% SDS, 0.75 µl proteinase K (1 mg/ml) and incubation for 20 min at 37 °C. The reaction was stopped by the addition of 20 µl 0.5 M EDTA.

2.4.2. IVT cRNA probe

2 µg of T7-amplified cRNA was labeled exactly as total RNA and polyA+ RNA by reverse transcribing and direct labeling with Cy3- or Cy5-dUTP except that random hexamers were used instead of oligo dT as primers.

2.4.3. SMART-cDNA probe

1 µg of PCR-amplified cDNA was labeled with Cy3- or Cy5-dCTP using random primers and Klenow enzyme. 1 µg of amplified cDNA was resuspended in 39 µl of 50 mM Tris (pH 6.8), 5 mM MgCl₂, 10 mM 2-mercaptoethanol and 300 µg/ml random nanomers. The mixture was incubated at 95 °C for 5 min and chilled on ice. After addition of 5 µl of dNTP mix (1.2 mM each dATP, dGTP, and dTTP; 0.6 mM dCTP), 3 µl of 1 mM Cy3- or Cy5-dUTP (Amersham Biosciences), and 3 µl (5 U/µl) Klenow fragment (New England Biolabs, Beverly, MA), the reaction mixture was incubated at 37 °C for 2 h and stopped with 5 µl of 0.5 M EDTA.

2.4.4. Probe purification and hybridization

For all labeled probes, each reaction was adjusted to 500 µl with 450 µl TE buffer (10 mM Tris, 1 mM EDTA), loaded onto a separate Microcon™ YM-30 filter (Millipore, Bedford, MA) and centrifuged at 8000 × g for 10 min. This washing step was repeated once more with another 450 µl of TE buffer. The labeled cDNA was recovered by placing the inverted Microcon™ filter in a new tube and spinning for 3 min at 5000 rpm (recovery was about 20–40 µl). Cy3- and Cy5-labeled probes were pooled together with 10 µg of mouse DNA Cot-1 (Invitrogen), 10 µg yeast tRNA and 20 µg poly (dA–dT) (Sigma, St. Louis, MO) and 10 µg poly (dA) (MWG Biotech, High Point, NC) and adjusted to a final volume of 24 µl in 3 × SSC, 0.2% SDS. The reaction mix was denatured at 95 °C for 3 min

prior to hybridization to a glass slide microarray. The array was covered by a glass cover slip, and the slide was placed in an individual humid chamber (Corning, New York, NY). Hybridization was carried out at 65 °C in a waterbath for 16–18 h. Post-hybridization washes were performed once with 2 × SSC, 0.2% SDS at 37 °C and once with 0.2 × SSC at room temperature for 5 min each.

2.5. Microarray construction

Mouse cDNA microarrays were constructed in the UCLA Microarray Core Facility (<http://www.genetic-s.ucla.edu/microarray/>). The inserts of 1920 non-redundant sequence verified mouse cDNA clones were amplified by PCR using vector specific M13 primers. PCR products were ethanol precipitated, washed in 70% ethanol, air-dried and resuspended in 300 mM sodium bicarbonate (pH 9.0) for printing on poly-L-lysine coated glass slides as previously described [22].

2.6. Image acquisition and data analysis

Hybridized arrays were scanned at a resolution of 10 µm utilizing a GMS 418 epi-confocal scanner (Affymetrix, Palo Alto, CA) to detect Cy3 and Cy5 fluorescence. Images of Cy3 and Cy5 signals were recorded separately as TIF-files and analyzed using SCANALYZE, version 2.44 [23]. cDNA spots that could not be interpreted due to artifacts introduced by slide processing were flagged and removed from further analysis. The distribution of the signal ratios Cy5/Cy3 for each type of hybridized probe was analyzed using GENESIGHT 3.0 (BioDiscovery, Los Angeles, CA). For all data, the local background was subtracted from the intensity value of each element on the array, and the negative values were removed. The data were converted to their log₂ values and centered using their Z values ($Z = [x - \mu_x] / \sigma_x$). To assess the reproducibility of the signal ratios Cy5/Cy3 of the replicate experiments, correlation coefficients between the experiments with each probe type were calculated in MICROSOFT EXCEL 2000 and the average was reported as R.

2.7. Northern analysis

Total RNA (30 µg) obtained from mouse liver and kidney were electrophoresed in 1.2% denaturing formaldehyde agarose gels. Following overnight capillary transfer to Hybond N+ membranes (Amersham, Arlington Heights, IL), RNAs were cross-linked to the membrane with the use of a Stratalinker (Stratagene, San Diego, CA). Probes were generated from cDNA plasmids using random primer labeling (Invitrogen) and ³²P-dCTP. Hybridizations were conducted at 68 °C

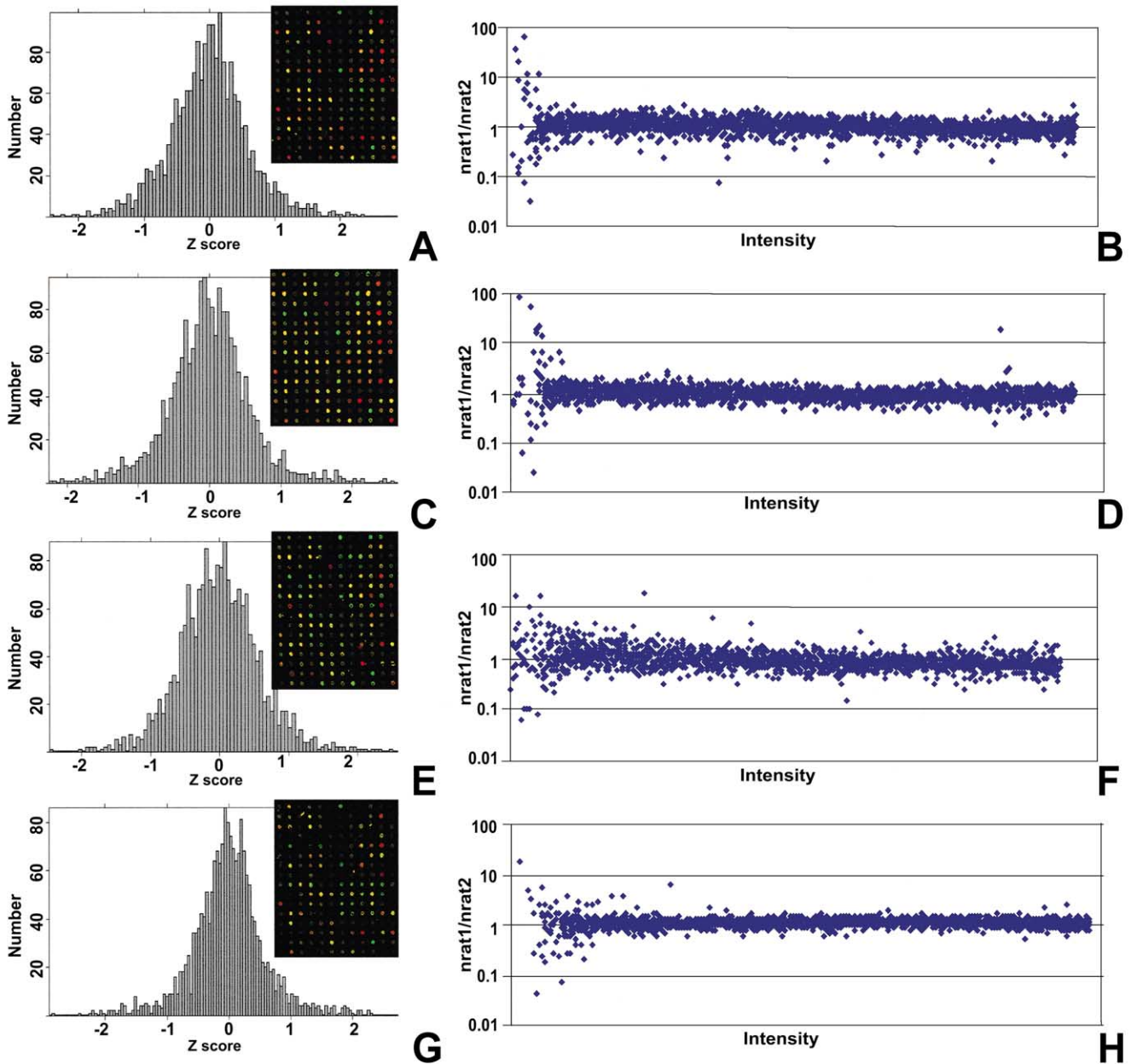


Fig. 1. Comparison of reproducibility of probe hybridization. Probes produced from RNAs of C57Bl/6 mouse liver and kidney using four different techniques were compared: (A–B) polyA+ RNA; (C–D) total RNA; (E–F) SMART-cDNA; (G–H) IVT cRNA. A, C, E and G are histograms representing the distribution of normalized expression ratios (Cy5/Cy3) for the set of 1920 genes hybridized by the different probes. The same array quadrant is shown after hybridization with each of the different probes; liver RNA was labeled with Cy5 (red color) and kidney RNA with Cy3 (green color). Genes that are transcribed equally in both tissues appear yellow. B, D, F and H are scatter plots representing the normalized ratios of one replicate versus another replicate ($nratio1/nratio2$) plotted against the intensity of the sum of the Cy3 and Cy5 signals.

using Express Hybridization Solution (Clontech). Post-hybridization washes were performed twice in $2 \times$ SSC, 0.1% SDS at room temperature, twice in $0.2 \times$ SSC, 0.1% SDS at 42°C and a final wash in $2 \times$ SSC at room temperature. Hybridized blots were autoradiographed on X-ray film.

3. Results

To limit the variability of the four different probe types, total RNA from the same extraction was used for polyA+ RNA purification and as the template for SMART and IVT amplifications. To generate probe for

Table 1
Consistency of the results obtained using different RNA probes

Labeled probe	Correlation coefficient of all intensity values of triplicate experiments	Correlation coefficient of top 75% intensity values of triplicate experiments
polyA+ RNA	0.70	0.94
Total RNA	0.54	0.91
SMART-cDNA	0.52	0.91
IVT cRNA	0.56	0.89

comparison to non-amplified procedures, mouse liver and kidney total RNAs (1 μg each) were amplified using cDNA amplification with the SMART system as well as cRNA amplification based on IVT with T7-RNA polymerase. Each amplification method was performed separately three times, and the resulting products were hybridized individually. To generate sufficient SMART-cDNA for each probe, 1/50th of the initial reverse transcription reaction was amplified in four different PCR reactions for 15 cycles (1 μl of reverse transcribed cDNA for each reaction), which is the optimized cycle in order to keep the large variances in the rate of amplification of different transcripts. Optimization in practice is determined as the maximum number of cycles until one tenth of the PCR product is visible as a smear on ethidium bromide stained gel. Amplified cDNAs were pooled and concentrated which resulted in yields of about 2 μg of amplified double-stranded cDNA. One microgram of SMART-cDNA probe was needed for labeling with Klenow. For the IVT cRNA probes, 1 μg of total RNA generated yields of 12–18 μg of amplified cRNA. Two micrograms of each IVT cRNA were reverse transcribed with random hexamers to generate a labeled sense cDNA probe. Two micrograms of mRNA were used as template to generate the polyA+ RNA labeled probes and 70 μg of total RNA were used as template to generate the total RNA labeled probes. All probes were labeled by direct incorporation of Cyanine-3 (Cy3) and Cyanine-5 (Cy5)-labeled nucleotide analogues. The amounts of template used for labeling were selected to result in similar total fluorescence incorporation into the probes. All probes were assessed by hybridization to arrays containing 1920 mouse cDNA clones deposited on poly-L-lysine coated glass slides.

3.1. Reproducibility

Each hybridization was compared with the other replicates from the same probe type to assess reproducibility. Fig. 1(left panel) demonstrates false colored red/green images of a portion of the array, which shows very similar relative intensities at the cDNA spots for each of the probes utilized. To visualize the reproducibility of the hybridization replicates within each group, we

plotted the quotient of the normalized expression ratio Cy5/Cy3 (nrat) of the genes in two replicates versus the intensity (as the sum of intensities in the Cy3 and Cy5 channel) of the genes (Fig. 1, right panel). A representative comparison graph is shown which indicates the reproducibility of expression ratios between two of the replicates of each probe over the entire range of signal intensities. In general, the signals with low intensity values (near background) that corresponded to transcripts with marginally detectable or non-detectable expression showed a higher variation in normalized expression ratios (nrat of Cy5/Cy3) within the replicates of all labeling methods (Fig. 1B, D, F, H). To quantitate the degree of reproducibility, we calculated the mean of the correlation coefficients (R) for the three replicates from each of the probe types. For the top 75% intensity values the correlation coefficient ranged from 0.94 to 0.89 (Table 1). In contrast, the bottom 25% intensity values (near background) showed a lower correlation for all probe types assayed.

3.2. Reliability of amplified probes

In order to assess the reliability of each type of probe to detect a large group of genes with modest differential expression, genes with a ratio value greater than 1Z or less than $-1Z$ were selected from each hybridization as differentially expressed from the mean of the polyA+ RNA experiments. The differential signals were on average about four-fold for this group of 382 genes. The same selected group of genes was re-plotted using the data obtained when the other three probe types were used for hybridization (Fig. 2). Within the new distribution plots, we considered those genes with a Z value between -0.5 and -1 or 0.5 and 1 as genes that were potentially in agreement with the less than $-1Z$ score and greater than 1Z score genes, respectively, selected from the experiment with polyA+ RNA probe. The genes with Z values between -0.5 and 0.5 were considered as false negatives. Table 2 shows that the percentage of total agreement among the probes ranged from 49 to 54%; the percentage of potential agreements among the probes was 65–75% and the percentage of no agreement (false negatives) between the alternate label-

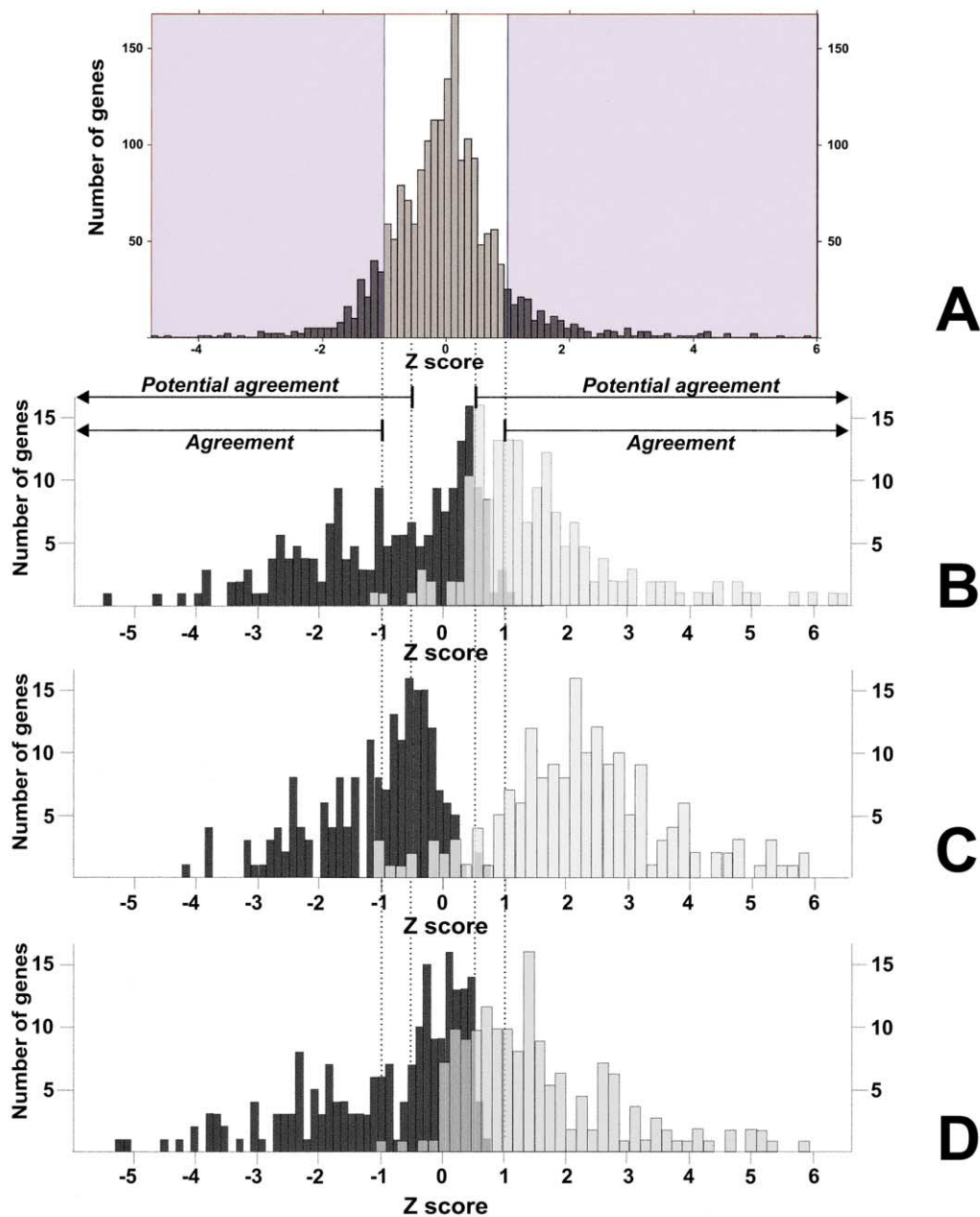


Fig. 2. Reliability of differential expression of each probe as compared with the polyA+ RNA probe. (A) Histogram representing the distribution of normalized expression ratios (Cy5/Cy3) for the set of 1920 genes hybridized by the polyA+ RNA probe. The shaded area indicates the genes with greater than $1Z$ or less than $-1Z$ value selected for comparison of hybridization with other probes. The same group of 382 genes is re-plotted using the mean signals obtained when the other probes were hybridized to the microarrays. Greater than $1Z$ values are re-plotted as light gray bars and less than $-1Z$ values as dark gray bars: (B) total RNA; (C) SMART-cDNA; (D) IVT cRNA.

ing methods and polyA+ RNA ranged from 19 to 35%. By this measure, the SMART-cDNA probe had the most similar distribution of genes detected as differentially expressed as the polyA+ RNA probe. However, many of the genes in the alternate probe labelings that were between $0.5Z$ and $-0.5Z$ were still differentially expressed as determined by the frequently used criteria of at least two-fold differentially expressed (Fig. 3).

Next we compared the strongly differentially expressed genes detected with each of the probes to determine how reliable gene expression detection was with any one probe relative to polyA+ RNA probe. Genes with ratios that were greater than $2Z$ or less than $-2Z$ values were selected from the mean of the three polyA+ RNA hybridizations (Table 3). With the polyA+ RNA probe, 93 genes were identified as

Table 2
Comparison of the sensitivity of probes to detect moderately differentially expressed genes

Probes	Total agreement ^a (%)	Potential agreement ^b (%)	No agreement ^c (%)
polyA + RNA	100		
Total RNA	52	74	26
SMART-cDNA	54	75	25
IVT cRNA	49	65	35

^a Percentage of 382 genes detected by polyA + RNA as $> 1Z$ or $< -1Z$ values from the mean of nrat (Cy5/Cy3) ratios that are also detected with $> 1Z$ or $< -1Z$ with alternative probes.

^b Percentage of 382 genes detected by polyA + RNA as $> 1Z$ or $< -1Z$ values from the mean of nrat (Cy5/Cy3) ratios that are also detected with $> 0.5Z$ or $< -0.5Z$ with alternative probes.

^c Percentage of 382 genes detected by polyA + RNA as $> Z$ or $< -Z$ values from the mean of nrat (Cy5/Cy3) ratios that have Z value from -0.5 to 0.5 with alternative probes.

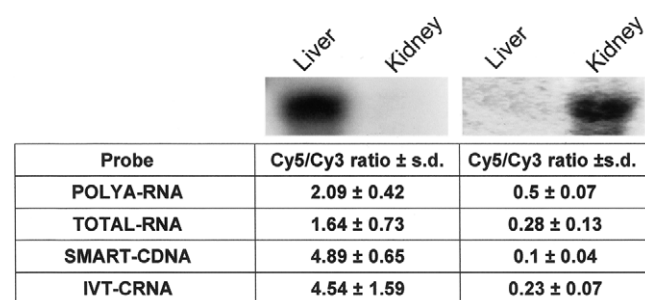


Fig. 3. Northern blot analysis of microarray detected differential expression. Northern blots confirm the differential expression of liver (Cy5) and kidney (Cy3) cDNAs detected by the microarrays using amplified probes. Two of the clones identified with amplified probes as differentially expressed were selected for hybridization onto Northern blots. Hybridization on the blot is depicted in the top row. The Cy5/Cy3-ratios of the microarray elements that hybridized to the different probes are shown below each Northern.

strongly differentially expressed between liver and kidney. There was an average eight-fold difference in expression for these 93 genes. We analyzed the results obtained in the hybridizations with the alternate probes for agreement with the polyA + RNA probe at these 93 genes. First each method was compared based on the greater than $2Z$ and less than $-2Z$ values of the ratios from each probe type. On this basis, the total RNA probe detected 73 of those 93 genes (78%) while the SMART-cDNA and IVT cRNA probes detected 61 and 65%, respectively. However, even when the number of genes detected by these probes was smaller than that detected by the polyA + RNA probe, their average Cy5/Cy3 signal ratio for the genes with Z values from $-2Z$ to $+2Z$ still showed as differential expression with a mean fold difference of about 3. If we take a more conservative measurement of concordance with the polyA + RNA probe of at least a two-fold difference in signal ratio, then the total RNA probe detects 99%, the SMART-cDNA probe detects 88%, and the IVT

cRNA probe detects 92% of the genes detected by the polyA + RNA probe (Table 3).

Each probe identified some differentially expressed genes that were not present in the greater than $2Z$ or less than $-2Z$ groups of any other probe (Table 3). For example, eight unique genes were detected with polyA + RNA and a different set of eight unique genes was observed with total RNA probes. A larger number of unique genes were identified by the amplified probes, but six of these were observed with both amplified probes. Northern blots of liver and kidney RNAs hybridized with two of the six genes indicated that these genes were indeed strongly differentially expressed and not an artifact of amplification (Fig. 3). Both genes were detected as being differentially expressed by the non-amplified probes but the degree of differential expression was underestimated. Northern blots of the other four genes gave low signals which were either too weak to detect their differential expressions or were nonspecific signals. In this verification, two of the six genes were confirmed being differential. This suggests that some genes uniquely detected in the amplified probes are indeed differentially expressed and would not have been detected in the unamplified probes. Thus, in some instances, it may be preferable to use an amplified probe to detect differential expression.

While there was generally strong agreement between the probe types, we noted a unique feature of the IVT cRNA probe which failed to detect significant signal at about 1% of the cDNA spots that were well detected with the polyA + RNA, total RNA, and SMART-cDNA probes. Presumably, these transcripts may have been poorly amplified in the IVT reactions.

4. Discussion

Gene expression profiling with DNA microarrays is a powerful technique being applied by an increasing number of investigators to study cellular pathways,

Table 3
Comparison of the sensitivity of probes for detecting strongly differentially expressed genes (greater than 2Z or less than -2Z)

Probe	Total number of detected genes	Genes from the 93 detected by polyA + RNA that were observed with each of the alternate probes	Genes with signal ratios > 2.0 or < 0.5 from the 93 detected by polyA + RNA that were observed with each of the alternate probes	Genes detected by any of two probes	Genes not detected by any other probes
polyA + RNA	93	93 (100%)	93 (100%)	85 (91%)	8 (9%)
Total RNA	93	73 (78%)	92 (99%)	85 (91%)	8 (9%)
SMART-cDNA	97	59 (61%)	82 (88%)	70 (72%)	27 (28%)
IVT cRNA	105	68 (65%)	86 (92%)	84 (80%)	21 (20%)

normal processes and diseases [3–7,9–16]. In many applications, the yield of cellular RNA extracted from the tissue being studied is sufficient to perform multiple hybridizations. However, in areas such as developmental biology, tumor diagnostics, or diseased biopsy analysis, investigators may need to study gene expression in only a few cells. In these cases, specimens obtained from microdissected tissues or needle biopsies contain a small amount of RNA that is not sufficient for microarray analysis [24,25]. This problem has been addressed by either amplifying the initial RNA (T7 and PCR amplification methods are currently being used) or enhancing the post-hybridization signal. The present work is the first comprehensive report on the reproducibility and reliability of SMART- and T7-IVT amplified probes when compared with non-amplified probes for spotted cDNA microarray analysis. All probes were obtained using a common source of total RNA, and all were hybridized to consecutive replicate mouse cDNA arrays to minimize non-probe related variation. T7-based linear amplification is frequently used to generate probes for microarray experiments, and has generally proven to be consistent. For instance, most of the probes generated for hybridization to Affymetrix GeneChip arrays use T7 RNA polymerase-based on IVT to generate biotinylated cRNA. However, PCR-based amplification schemes like the SMART system have not been as widely adopted.

Our results indicate that SMART-cDNA probes are as reliable as T7-IVT probes for the detection of differentially expressed genes. There are several practical reasons why we favor PCR-generated probes. First, the mRNA is converted into relatively more stable cDNA in the initial step, easing concerns about degradation during probe handling, long term probe storage, and probe hybridization on the spotted arrays. Second, the degree of amplification is larger, from one microgram of total RNA we typically obtain about 25 µg of double-stranded cDNA from which we can perform about 25 hybridizations. The typical yield from T7 to IVT is about 15 µg. Two micrograms of T7-IVT probe is optimal for 24 µl hybridization volumes. Thus, sufficient probe is generated for seven hybridizations. Nevertheless, both methods can be applied to smaller starting amounts of total RNA.

Similar results have been reported comparing linear and exponential amplification using IVT and extended RT-PCR, respectively, on RNA samples purified from cell lines [26]. However, they used one tenth to one fourth of the starting RNA used in this study for the linear amplification (T7-IVT) comparison, which may have caused a stronger discrepancy between the outcome of hybridization of T7-IVT amplified versus non-amplified probes compared with the present study. On the other hand, similar endpoint results were obtained with our full-length exponential amplification method (SMART-

cDNA), which requires fewer steps than the extended RT-PCR method.

Each of the amplification methods has the possibility to either not amplify a subset of mRNAs or to differentially amplify specific messages in replicate amplifications. Reproducibility is perhaps most essential for gene expression profiling in that even if all transcripts are not amplified, and therefore not analyzed, in an amplification method, useful relative differential gene expression information can still be obtained for the set of genes amplified. The correlation coefficients obtained from replicate hybridizations with the different probes reveal a high level of reproducibility with each of the methods. The intrinsic deviation, introduced by labeling and hybridization variation, is similar for all the probes, including polyA+ RNA and total RNA. Thus, within a probe labeling method, the comparative results obtained from a SMART-cDNA or IVT cRNA probe should give consistent results across a series of specimen. However, some genes may be dropped out and not measured at all especially by the T7-IVT method (data not shown). Conversely, not all genes are well measured with the non-amplified probes. For instance, both genes detected by Northern blot to be strongly differentially expressed also showed large differential Cy5/Cy3 signals with both IVT cRNA and SMART-cDNA but not with polyA+ RNA or total RNA. Higher signal intensity spots are more accurately measured on microarrays. Since the cDNA spots hybridized more brightly with the amplified than non-amplified probes, the accuracy of the Cy5/Cy3 ratio measurements was more accurate with the amplified probes. Cy-dye conjugated dCTP was used to label SMART-cDNA and Cy-dye coupled dUTP for the generation of the other three types of probes. Since different labeling methods were employed for generation of different probes, differential Cy-dye incorporation based on sequence identity of the template may occur.

There are discrepancies in measured gene expression between the two non-amplified probes (polyA+ RNA and total RNA). The polyA+ probe is purified from the original total RNA pool. Thus, the polyA+ probe is presumed to be less contaminated with ribosomal and transfer RNA that may interfere during hybridization and get spuriously labeled and create cross-hybridization. On the other hand, the polyA+ probe may contain only the higher abundant RNA species as a purification effect, and therefore, produces a 'cleaner' expression profile. It is interesting to note that the choice of using purified mRNA as the starting material for labeling as opposed to total RNA caused as many differences as the amplification of the probes.

As we observed high reproducibility in the results obtained with each probe and consistency in the expression ratios of the 1920 murine cDNAs expressed in liver or kidney, we conclude that the relative

abundance of RNA species in the amplified probes is largely maintained during the amplification process. Thus, by using these amplified probes, it is possible to start with 1 µg of total RNA, an amount that is up to 100 times lower than conventionally used for microarray experiments, and generate sufficient probe for multiple replicate hybridizations.

Acknowledgements

We thank Barry Merriman for his advice in data analysis and the UCLA Microarray Core Facility for providing cDNA arrays. This work was supported by an administrative Supplement to support DNA Microarray Facilities Program No EY00-001 (DBF) and EY10836 (MCK) of the National Eye Institute, UO1 CA88127 (SFN) from the NCI and Guenther Foundation (DB). DBF is the recipient of a Research to Prevent Blindness Senior Research Investigators Award.

References

- [1] Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. *Science* 1995;270:467–70.
- [2] Endege WO, Steinmann KE, Boardman LA, Thibodeau SN, Schlegel R. *Biotechniques* 1999;26:542–8.
- [3] Spirin KS, Ljubimov AV, Castellon R, Wiedoeft O, Marano M, Sheppard D, Kenney MC, Brown DJ. *Invest Ophthalmol Vis Sci* 1999;40:3108–15.
- [4] Gonzales P, Epstein DL, Borrás T. *Invest Ophthalmol Vis Sci* 2000;41:352–9.
- [5] Zhumbayeva B, Diatschenko L, Chenchik A, Siebert PD. *Biotechniques* 2001;30:158–63.
- [6] Herrler M. *J Mol Med* 2000;78:B23.
- [7] Livesey FJ, Furukawa T, Steffen MA, Church GM, Cepko CL. *Curr Biol* 2000;10:301–10.
- [8] Van Gelder RN, Von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. *Proc Natl Acad Sci USA* 1990;87:1663–7.
- [9] Dieckgraefe BK, Stenson WF, Korzenik JR, Swanson PE, Harrington CA. *Physiol Genomics* 2000;4:1–11.
- [10] Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM. *Nat Biotechnol* 2000;18:457–9.
- [11] Lee CK, Klopp RG, Weindruch R, Prolla TA. *Science* 1999;285:1390–3.
- [12] Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH, Levine AJ. *Genes Dev* 2000;14:981–93.
- [13] Voehringer DW, Hirschberg DL, Xiao J, Lu Q, Roederer M, Lock CB, Herzenberg LA, Steinman L, Herzenberg LA. *Proc Natl Acad Sci USA* 2000;97:2680–5.
- [14] Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJ, Morris D, Huang X, Sheppard D, Heller RA. *Proc Natl Acad Sci USA* 2000;97:1778–83.
- [15] Pabon C, Modrusan Z, Ruvolo MV, Coleman IM, Daniel S, Yue H, Arnold LJ, Jr. *Biotechniques* 2001;31:874–9.
- [16] Ohyama H, Zhang X, Kohno Y, Alevizos I, Posner M, Wong DT, Todd R. *Biotechniques* 2000;29:530–6.
- [17] Luo GX, Taylor J. *J Virol* 1990;64:4321–8.

- [18] Mills GB, Lu Y, Fang X, Wang H, Eder A, Mao M, Swaby R, Cheng KW, Stokoe D, Siminovitch K, Jaffe R, Gray J. *Semin Oncol* 2001;28:125–41.
- [19] Mills JC, Roth KA, Cagan RL, Gordon JI. *Nat Cell Biol* 2001;3:E175–8.
- [20] Karsten SL, Van Deerlin VMD, Sabatti C, Gill L, Geschwind DH. *Nucleic Acids Res* 2002;30:E4.
- [21] Stears RL, Getts RC, Gullans SR. *Physiol Genomics* 2000;3:93–9.
- [22] Welford SM, Gregg J, Chen E, Garrison D, Sorensen PH, Denny CT, Nelson SF. *Nucleic Acids Res* 1998;26:3059–65.
- [23] Eisen, M. 1999; <http://www.rana.lbl.gov/EisenSoftware.htm>
- [24] Geschwind DH. *Proc Natl Acad Sci USA* 2000;97:10676–8.
- [25] Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR, Erlander MG. *Nat Med* 1999;5:117–22.
- [26] Iscove NN, Barbara M, Gu M, Modi C, Winegarden N. *Naturebiotechnology* 2002;20:940–3.