MICROARRAY Technologies

Research Report

Advantages of mRNA Amplification for Microarray Analysis

BioTechniques 33:906-914 (October 2002)

A.L. Feldman, N.G. Costouros, E. Wang[1], M. Qian, F.M. Marincola[1], H.R. Alexander, and S.K. Libutti
National Cancer Institute and [1]National Institutes of Health, Bethesda, MD, USA

ABSTRACT

Expanding applications of cDNA microarrays such as fine needle aspiration biopsy and laser capture microdissection necessitate the ability to perform arrays with minute starting amounts of RNA. While methods for amplifying RNA have been advocated, the fidelity of array results using amplified material has not been fully validated. Here we demonstrate preserved fidelity in arrays using one or two rounds of mRNA amplification, validated by downstream real-time quantitative PCR. In addition, the quality of the array data was superior to that obtained using total RNA. Based on these results, we recommend routine mRNA amplification for all cDNA microarray-based analysis of gene expression.

INTRODUCTION

Recent advances in microarray technology (1,3) coupled with the sequencing of human and other genomes (2,5) have enabled high-throughput gene expression profiling. This technology has led to powerful advances in identifying genetic profiles of human diseases as well as the ability to screen thousands of individual genes that may be expressed differentially between two samples (1). Glass slide cDNA microarray analysis requires relatively large amounts of total RNA (50–200 µg). However, some sources of RNA have limited yield, including fine needle aspirate samples (9) and samples that are isolated using laser capture microdissection (7,8,10). In addition, the quality of RNA can vary substantially, depending on the method of isolation, presence of endogenous RNases, and other factors that may be introduced in the clinical setting (e.g., time of tissue harvest) (4).

Techniques for mRNA amplification have been used to analyze gene expression from small amounts of total RNA (11,12). Initial data suggest that these methods can be used successfully for microarray analysis (7,8,12). However, the ability of this approach to identify accurately outlying genes has not been fully validated. We evaluated the quality of microarray data obtained using total and amplified mRNA and used real-time quantitative PCR to assess the validity of genes identified as outliers. Based on the improved quality, preserved fidelity, and ability to use much smaller amounts of source material using amplified mRNA, we recommend routine mRNA amplification for all cDNA microarray-based analysis of gene expression.

MATERIALS AND METHODS

Cell Lines and RNA Amplification

B16F10 murine melanoma cells (ATCC, Manassas, VA, USA) and MC38 murine colon adenocarcinoma cells [developed in the Surgery Branch, National Cancer Institute (NCI), Bethesda, MD, USA] were maintained in DMEM with 10% FCS, 100 U/mL penicillin, 100 µg/mL streptomycin, 50 µg/mL gentamicin, 0.5 µg/mL Fungizone®, and 4 mM glutamine (Biofluids, Rockville, MD, USA). Total RNA was isolated using TRIZOL® reagent (Invitrogen, Carlsbad, CA, USA) and further purified using RNeasy® mini kits (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions.

Amplified antisense RNA was prepared as previously described (11), with minor modifications. Briefly, 3 µg total RNA were incubated with 2 µg T7-$(dT)_{24}$ primer (5′-GGCCAGTGAATTGAATACGACTCACTATAGGGAGGCGG-$(dT)_{24}$-3′) in 9 µL DEPC-treated water at 70°C for 7 min and cooled on ice. cDNA was synthesized by the addition of 4 µL first-strand buffer, 2 µL 0.1 M DTT, 2 µL 10 mM dNTP, 2 µL SuperScript® II reverse transcriptase (Invitrogen), and 1 µL RNasin® (Promega, Madison, WI, USA) and incubation at 42°C for 1 h. Second-strand synthesis was performed by the addition of 91 µL DEPC-treated water, 30 µL second-strand buffer, 3 µL 10 mM dNTP, 4 µL

Table 1. Real-Time Quantitative PCR Primers and Probes

Gene	Forward Primer (5′→3′)	Reverse Primer (5′→3′)	Probe (5′→3′)
Gpi1 (NM_008155)	AGCACTTTGTCGCCC-TGTCTA	GAGCAATGGAAAGTCCAATGG	FAM-ACCACAGCGAATAGCG-GCCACCT-TAMRA
Thbs1 (NM_011580)	CAGGGCAAAGAACTC-GTCCA	CGAATACCAAGCCAGCGTAGT	FAM-ACAGCATTAAACTCATCAT-AACCTACAGCAAGTCCAG-TAMRA
Serpine1 (NM_008871)	TGGAGGGCACAACA-CTTTCA	GGACATTTCCACAGTGGACCTT	FAM-TCTTGCATCGCCTGCCATTG-CTG-TAMRA
Asml3a (NM_020561)	TTGTGGCACCTGCTG-TTACAC	TGCACCATGTCCAGCAATG	FAM-TCCGCCTATTTCAGTACAAGC-CTGGTGA-TAMRA
Prdx5 (NM_007453)	CATCTTGCCTGGAG-CAAGGA	GGATCCAACATGCAGAAAAGG	FAM-ATCAATGCTTACAATGGTG-CAACACCCA-TAMRA

NM numbers in parentheses represent UniGene codes; FAM-TAMRA sequences represent probes. C_t, threshold cycle number for detection.

E. coli DNA polymerase I, 1 µL *E. coli* DNA ligase, and 1 µL RNase H- (Invitrogen) and incubation at 16°C for 2 h. T4 DNA polymerase (2 µL; Invitrogen) was added, and the reaction was incubated at 16°C for 5 min. The reaction was stopped by the addition of 10 µL 0.5 M EDTA and 10 µL 1 M NaOH and incubation at 70°C for 10 min and then neutralized with 25 µL Tris-HCl, pH 7.5. DNA was extracted with phenol: chloroform: isoamyl alcohol and precipitated in ethanol with 5 µg linear acrylamide (Ambion, Austin, TX, USA). Transcription in vitro was carried out at 37°C for 5 h in a 40-µL reaction volume using the T7 Megascript™ kit (Ambion), according to the manufacturer's instructions. RNA was purified in an RNeasy mini column or phenol:chloroform: isoamyl alcohol extracted/ethanol-precipitated if proceeding to second-round amplification. Second-round amplification was performed similarly, with slight modifications. The amplified RNA was resuspended in 11 µL DEPC-treated water and primed with 1 µg random hexamers (Invitrogen). Before second-strand synthesis, cDNA was incubated with 1 µL RNase H- for 20 min at 37°C and then heated to 95°C for 2 min. Second-strand synthesis was then primed with 0.5 µg T7-$(dT)_{24}$ by incubation at 70°C for 5 min and 42°C for 10 min. The remaining steps were identical to the first-round amplification, and the amplified RNA was purified in an RNeasy mini column. Total and amplified RNA was analyzed for purity and concentration by formaldehyde gel electrophoresis and spectrophotometry. Similar experiments were performed to amplify and compare RNA from human cell lines (unpublished data) and rat aortic rings (13).

Total and amplified RNA was applied to the RNA 6000 Labchip® using a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA, USA). The chip was prepared according to the manufacturer's protocol. Briefly, the gel/dye mixture was prepared by centrifugation of 400 µL gel matrix and 4 µL RNA dye through a spin filter (supplied with the kit) at 2000× *g* for 10 min. The gel/dye mixture (9 µL) was used to fill the channels of the RNA chip. Sample buffer (5 µL) was added to each sample and molecular marker well. RNA 6000 ladder (1 µL or 150 ng; Ambion) or test RNA sample (1 µL, diluted 1:10) was added to each well. After vortex mixing for 1 min at 2000 rpm, an RNA smear assay was run on the chip, and the size of the rRNA in the total RNA sample and amplified RNA was compared with the molecular weight markers.

Microarray Analysis

Microarray analysis was conducted to compare gene expression in MC38 and B16F10 using total RNA or RNA after one or two rounds of amplification. Microarray probes were synthesized using 30 µg (for Cy3 labeling) or 50 µg (for Cy5 labeling) total RNA or 3 µg amplified RNA using 2 µg oligo-$(dT)_{20}$ or random hexamers, respectively, for priming. cDNA was generated using Cy3- or Cy5-dUTP (Amersham Biosciences, Piscataway, NJ, USA) and SuperScript II reverse transcriptase. Cy3- and Cy5-containing probes were combined and purified in Microcon® columns (YM-30; Millipore, Bedford, MA, USA) and denatured at 100°C for 1 min with 10 µg mouse COT-1 DNA (Invitrogen), 4 µg yeast tRNA (Sigma, St. Louis, MO, USA), and 8 µg poly(A) (Amersham Biosciences). NCI glass slide microarrays representing 2601 mouse genes were prehybridized at 42°C for 1 h in 5× SSC, 0.1% SDS, and 1% BSA. The arrays were then hybridized with a probe in 25% formamide, 5× SSC, and 0.1% SDS at 42°C overnight (10–16 h). The slides were then washed for 2 min in 2× SSC, 0.1% SDS; 2 min in 1× SSC, 0.1% SDS; 2 min in 0.2× SSC; and 1 min in 0.05× SSC and were then spun dry at 100× *g*. Fluorescence images were captured using a Genepix® 4000 (Axon Instruments, Foster City, CA, USA). Data were stored in a database supported by the Center for Information Technology, National Institutes of Health and analyzed using software developed at the National Human Genome Research Institute (Bethesda, MD, USA). The ratios were expressed as the relative gene expression in B16F10: MC38. To establish reproducibility and minimize the effects of labeling bias, all arrays were repeated using reciprocal fluorescence (i.e., RNA targets labeled with Cy3 in the first array were labeled with Cy5 in the second array).

An outlier for a given RNA sample (total or amplified) was defined as a gene with expression ratios of ≥ 2.0 or ≤ 0.5 in two reciprocal microarray analyses (ratios ≥ 2.0 were ≤ 0.5 when the Cy3 and Cy5 fluorochromes were reversed and vice versa); fluorescence intensities were ≥ 100 (based on a scale from 1 to 65 536 U), and spot sizes were ≥ 50 pixels. Genes identified as outliers by microarray analysis were categorized according to the following scheme: category I, outlier using total RNA but not amplified RNA; category II, outlier using both total RNA and amplified RNA; category III, outlier using amplified RNA with a similar trend using total RNA (showing expression ratios in the same direction as with amplified RNA, but not meeting the threshold value of 2.0 on both arrays); category IV, outlier using amplified RNA with insufficient data using total RNA (caused by uninterpretable array spots that did not meet filtering criteria); and category V, outlier using amplified RNA with discordant results using total RNA (in which either array yielded a ratio in the opposite direction as that with amplified RNA). Mean ratios were calculated using log ratios and reciprocal ratios for downregulated genes.

Real-Time Quantitative PCR

Real-time quantitative PCR analysis was carried out as described previously (6), with minor modifications. Briefly, measurements were performed using the GeneAmp® 5700 Sequence Detector (Applied Biosystems, Foster City, CA, USA). The primers and probes (BioServe Biotechnologies, Laurel, MD, USA) were designed using Primer Express software (Applied Biosystems), except for rodent GAPDH, which was available commercially (Applied Biosystems). Probes were 5′-labeled with 6-carboxyfluorescein (FAM) and 3′-labeled with 6-carboxytetramethylrhodamine (TAMRA). Table 1 shows these sequences.

cDNA was generated from RNA samples (50 ng/assay for total RNA and 2 ng/assay for amplified RNA) using MultiScribe® reverse transcriptase and random hexamers for priming (Applied Biosystems). Reaction conditions were at 25°C for 10 min, 48°C for 30 min,

Table 2. Calculation of Copy Numbers for Sample Gene (Thbs1) Based on Extrapolation from Standard Curve

Generation of Standard Curve (Thbs1)

Copy No.	C_t (mean)
1×10^8	16.6
1×10^6	23.2
1×10^4	31.3
0	>40 (∞)

Unknown Samples (Thbs1)

Sample	C_t (mean)	Extrapolated Copy No. (mean)
MC38, total RNA	22.2	2.5×10^6
MC38, 1-round amp.	24.9	4.8×10^5
MC38, 2-round amp.	28.0	6.6×10^4
B16F10, total RNA	31.7	6.5×10^3
B16F10, 1-round amp.	34.2	2.9×10^3
B16F10, 2-round amp.	39.1	6.7×10^1

Standard curves and extrapolated copy numbers were generated by GeneAmp 5700 Sequence Detector software. The samples were normalized to the GAPDH copy number calculated in the same manner. amp., amplification.

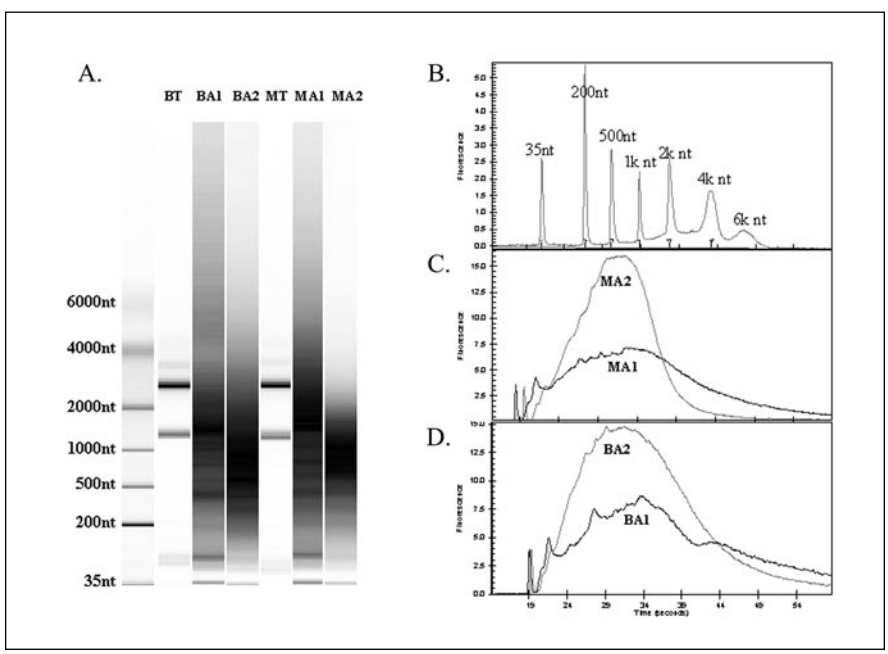

Figure 1. Analysis of RNA quality and length. (A) RNA gel images were generated using a Bioanalyzer 2100 and an RNA 6000 chip on an RNA smear assay. The size of molecular weight markers is indicated according to the analysis program. 28S and 18S rRNA bands from the total RNA samples are demonstrated in lanes BT and MT. Second-round amplified RNA (lanes BA2 and MA2) shows truncation in size compared to first-round amplified RNA (lanes BA1 and MA1). Molecular markers (B) were compared to electropherograms of the amplified samples (C, cell line MC38; D, cell line B16F10) to determine the size range of the first-round and second-round amplified RNA using the overlay function in the bioanalyzer software. Most of the first-round amplified RNA transcripts (A1) ranged from 200 to 4000 nucleotides, while most of the second-round amplified RNA transcripts (A2) ranged from 100 to 2000 nucleotides. nt, nucleotides.

and 95°C for 5 min. cDNA standards for each gene were generated by primer-specific amplification of MC38 and B16F10 cDNA mixed in equal parts and analyzed for purity and concentration by gel electrophoresis and spectrophotometry. Copies were calculated using the molecular weight of each product. The real-time PCR reactions for samples and standards were conducted in a volume of 25 µL using AmpliTaq® Gold DNA polymerase (Applied Biosystems), sense and antisense primer concentrations of 900 nM each, and a probe concentration of 250 nM (except for GAPDH, for which primer and probe concentrations each were 100 nM, according to the manufacturer's recommendations). Thermal cycler parameters were at 50°C for 2 min, 95°C for 10 min, 40 cycles of a 15-s denaturing step at 95°C, and a 1-min annealing/extension step at 60°C. Standard curves were generated for each gene, and copy numbers were extrapolated for each sample (Table 2). All assays were performed in duplicate and reported as the mean. The relative expression of each gene of interest was normalized to that of GAPDH, and the expression ratio for each gene in B16F10:MC38 was calculated.

RESULTS

Characterization and Quality of RNA

RNA quality and length were measured using the RNA smear assay (Figure 1A) and electropherograms of the amplified samples (Figure 1, B–D). The starting total RNA demonstrated clear 28S and 18S rRNA bands; the RNA smears of the amplified samples demonstrated truncation of second-round amplified RNA compared to first-round amplified RNA. Most transcripts were in the 200–4000 nucleotide range after one round of amplification and in the 100–2000 nucleotide range after two rounds.

The correlation coefficient for the reciprocal arrays using total RNA was -0.67 (Figure 2A), which was similar to that obtained with reciprocal analyses in other experiments using total RNA (data not shown). When amplified RNA was used, correlation coefficients ranged from -0.92 after one round of amplification (Figure 2B) to -0.94 after two rounds (Figure 2C). Using total RNA, the correlation coefficients between duplicate forward arrays or duplicate reciprocal arrays ranged from 0.72 to 0.80. Correlation coefficients between arrays after one round of amplification and arrays after two rounds of amplification ranged from 0.93 to 0.95. Correlation coefficients between arrays using total RNA and arrays using amplified RNA ranged from 0.77 to 0.82 after one round of amplification and from 0.75 to 0.81 after two rounds of amplification.

Outlier Identification and Validation

One hundred thirty-one genes were identified as outliers using total RNA, 239 were identified using RNA after one round of amplification, and 257 were identified using RNA after two rounds of amplification (Figure 3). Analysis of amplified RNA identified 81% (Figure 3A; one round of amplification) or 79% (Figure 3B; two rounds of amplification) of genes identified using total RNA. The concordance between outliers identified after one and two rounds of amplification was from 77% to 82% (Figure 3C; 197 genes in common).

Among genes identified as outliers using amplified RNA, 40%–44% of the genes were concordant (category II) based on total RNA array results; 38%–41% of the genes showed a trend toward concordance (category III); 12%–18% of the genes demonstrated insufficient data (category IV); and only 3%–4% of the genes were discordant (category V) (Figure 4A). Twenty-eight percent of the genes identified as outliers using single-round amplified RNA exhibited missing data points on total RNA arrays. The mean ratio on the total RNA arrays of the genes identified as outliers using amplified RNA but not total RNA was 2.01. Similar results were

obtained comparing total RNA and RNA after two rounds of amplification.

The expression of nine genes in total and amplified RNA was analyzed using real-time quantitative PCR (Figure 4, B–F). Genes from categories I–IV (Figure 4, B–E) were validated by real-time quantitative PCR as being expressed differentially in all samples (total, one round of amplification, and two rounds of amplification), with microarray analysis generally underestimating the quantitatively determined expression ratio. In category V (Figure 4F), real-time quantitative PCR data were discordant among the samples and did not confirm the true differential expression of these genes. Similar results were obtained with other genes from these categories (data not shown).

DISCUSSION

As the use of microarray technology became more widespread, applications that use minute amounts of starting material emerged. These applications include minimally invasive procedures to obtain clinical material for diagnostic and research use, such as fine needle aspiration (9), as well as advanced technology to isolate highly specific tissue samples ex vivo, including laser capture microdissection (7,8,10). This study demonstrates that amplification of RNA samples for microarray analysis is a highly sensitive method to identify outlier genes that retains its specificity when compared to techniques using total unamplified RNA. These findings were supported by the results of downstream real-time quantitative PCR. In addition, the quality of microarray data obtained using amplified mRNA was superior to that obtained using total RNA.

We utilized an antisense amplification technique that was expected to produce truncation of the transcript length with successive rounds of amplification. This truncation was observed by RNA smear assay and electropherograms of the amplified samples, with the transcript size reduced by approximately 50% with the second round of amplification. The quality and reproducibility of microarray data generated using total or amplified RNA were evaluated in part by assessing the correlation between expression ratios obtained in two independent, reciprocal arrays. Reciprocal arrays (reversing the fluorescent labels used during cDNA synthesis) were analyzed to evaluate possible effects of labeling bias. In this study, we utilized RNA from standard murine tumor cell lines (B16F10 and MC38) to ensure sufficient total RNA for the purpose of meaningful comparison with amplified RNA. Correlation coefficients were consistently higher when amplified RNA was used and were preserved with a subsequent round of amplification. Using total RNA, the correlation between duplicate forward arrays or duplicate reciprocal arrays was slightly stronger than the correlation between forward and reciprocal arrays. Cy3/Cy5-labeling bias may play a role in the lower correlation between ratios using total RNA but does not fully explain the difference in the degree of correlation. Increased interarray variability also appears present using total RNA. Signals corresponding to low-copy transcripts in total RNA samples may be more affected by background noise in repeat arrays and thus would be less likely to correlate strongly than the enriched, high-quality transcripts in amplified material. The correlation between the arrays after one round of amplification and the arrays after two rounds of amplification remained similar, which suggests that no substantial bias was introduced by the second round of amplification. The correlation between the arrays using total RNA and the arrays using amplified RNA was similar to the correlation between duplicate arrays using total RNA, providing further evidence of a lack of bias introduced by amplification.

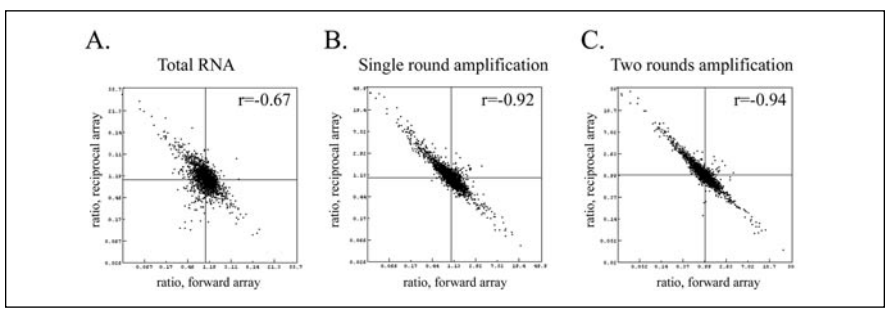

Figure 2. Correlations in reciprocal microarray experiments using total and amplified RNA. Correlations between gene expression ratios obtained in two independent microarray analyses are shown using total RNA (A) or RNA after one (B) or two (C) rounds of antisense amplification. Ratios in forward arrays (x-axes) represent relative gene expression in B16F10 RNA labeled with Cy5 compared to MC38 RNA labeled with Cy3. For reciprocal arrays (y-axes), each RNA target was labeled with the opposite fluorochrome. The correlation coefficients (r) for reciprocal arrays performed using RNA after one (r = -0.92) or two (r = -0.94) rounds of amplification were higher than for arrays performed using total RNA (r = -67).

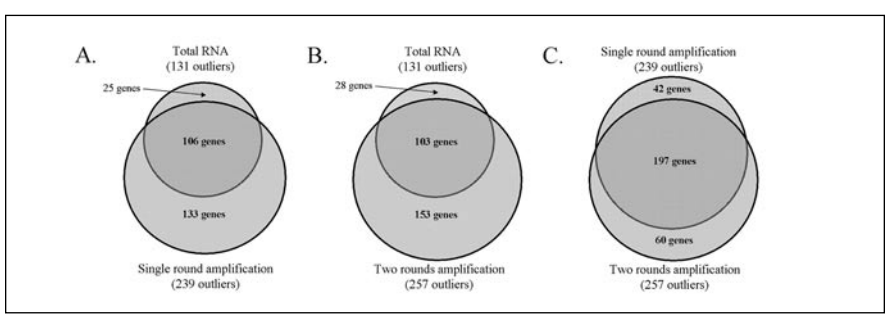

Figure 3. Venn diagrams comparing genes identified as outliers using total and amplified RNA. The Venn diagrams represent genes identified as outliers based on expression ratios ≥ 2.0 or ≤ 0.5 in reciprocal microarray analyses (ratios ≥ 2.0 were ≤ 0.5 when the Cy3 and Cy5 fluorochromes were reversed and vice versa). Analysis of amplified RNA identified 81% (A, one round of amplification) or 79% (B, two rounds of amplification) of genes identified using total RNA. The number of outliers identified after one round of amplification did not increase substantially after a second round of amplification (C), and the concordance was from 77% to 82%. However, the analysis of the amplified RNA yielded 1.8–2.0 times more outliers than total RNA, which yielded a substantial pool of genes identified by analysis of amplified RNA but not total RNA.

These data suggest that multi-gene molecular profiles characterized by microarray analysis may be preserved after RNA amplification. However, it is unclear whether individual genes that are identified as being expressed differentially using amplified RNA are likely to be genes that would have been identified as such using total RNA. To answer this question, we compared outliers that had been identified using total RNA with those identified using RNA after one or two rounds of amplification. Using criteria described above, 131 genes were identified as outliers using total RNA, 239 outliers were identified using RNA after one round of amplification, and 257 outliers were identified using RNA after two rounds of amplification. Analysis of amplified RNA identified approximately 80% of the genes identified using total RNA, and the concordance between outliers identified after one and two rounds of amplification was also approximately 80%. It is likely that the strong but not perfect correlations between different RNA amplifications (i.e., $r < 1.0$), combined with the stringency criteria set for our definition of an outlier, explain the lack of complete concordance in outliers identified between different RNA amplifications. Because analysis of amplified RNA yielded 1.8–2.0 times more outliers than total RNA, a substantial pool of outliers was identified by the analysis of amplified RNA but not total RNA (up to 60% of the outliers identified using amplified RNA). Thus, the validity of identifying these genes as being expressed differentially is critical to the use of amplified RNA in microarray analysis. It is possible that the improved quality of microarray data obtained using amplified RNA allows a greater number of genes to meet the criteria for the definition of an outlier than data using total RNA. However, it is also possible that this large pool of up to 60% of identified outliers represents genes whose amplification is not linear or is otherwise biased, rendering these outliers false positives that would seriously flaw gene expression profiles generated from amplified samples and hinder downstream validation and interpretation of individual genes of potential scientific interest.

To address this important issue, we evaluated expression patterns of each gene identified as an outlier in any RNA sample (total or amplified). If the differences between the outliers identified by total and amplified RNA were caused by amplification bias, then we would expect substantial differences in expression patterns in total and amplified RNA among those genes not identified as outliers using both RNA sources. Thus, the mean ratio of genes identified by amplified RNA but not total RNA (or vice versa) would be

expected to be close to 1.0 (to which the mean ratio for all genes on the array is normalized). On the other hand, if differences between outliers identified by total and amplified RNA were caused by usual interarray variation and a greater sensitivity of the amplified RNA arrays to detect outliers, then we would expect (*i*) trends for the genes identified by amplified RNA but "missed" by total RNA to have ratios with the same direction of change, even if the magnitude did not reach ≥ 2.0 or ≤ 0.5; (*ii*) a substantial number of missing data points among the genes identi-

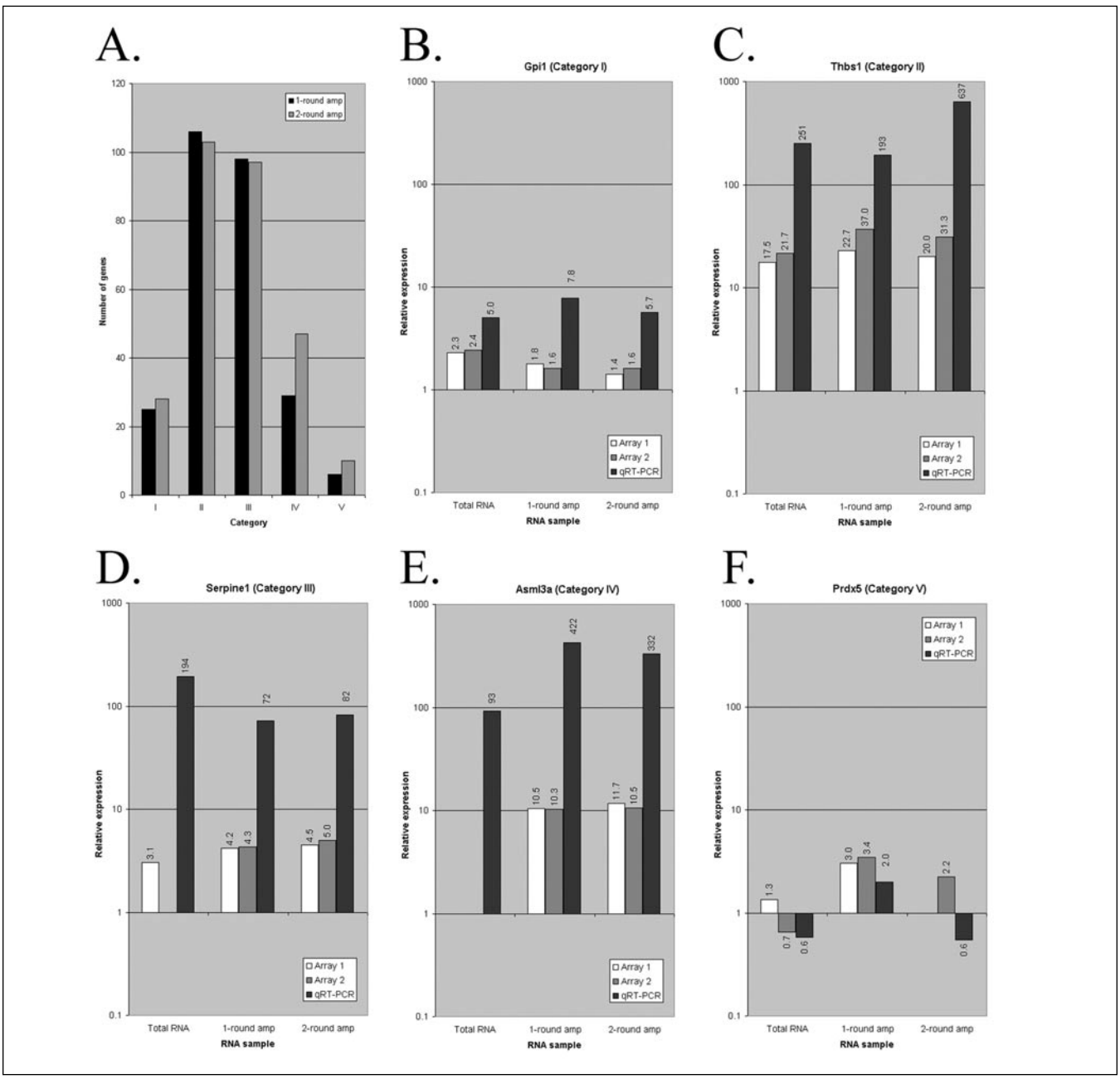

Figure 4. Categorization and validation of genes identified as outliers by microarray analysis. (A) Genes identified as outliers by microarray analysis were categorized according to the following scheme: category I, outlier using total RNA but not amplified RNA; category II, outlier using both total RNA and amplified RNA; category III, outlier using amplified RNA with similar trend using total RNA; category IV, outlier using amplified RNA with insufficient data using total RNA; and category V, outlier using amplified RNA with discordant results using total RNA. After one or two rounds of amplification, the majority of genes were in categories II or III, with only 3%–4% of genes being discordant (category V). (B–F) The expression ratios of the genes in each category were evaluated further using real-time quantitative PCR. Missing bars represent spots that did not meet filtering criteria on the arrays so that expression ratios could not be characterized. Genes from categories I–IV were validated as being expressed differentially in all samples (total, one round of amplification, and two rounds of amplification), with microarray analysis generally underestimating the quantitatively determined expression ratio. In category V, real-time quantitative PCR data were discordant among the samples and did not confirm the true differential expression of these genes. Similar results were obtained with other genes from these categories (data not shown). qRT-PCR, real-time quantitative PCR.

fied by amplified RNA but not total RNA; and (*iii*) a mean ratio for these genes on total RNA arrays that is closer to the threshold value of 2.0 rather than the normalized value of 1.0. Genes identified as outliers by microarray analysis were categorized according to a scheme (described earlier) designed to group genes by the concordance of the array results using total and amplified RNA. Based on this scheme, approximately 80% of the genes identified as outliers using amplified RNA were either concordant (category II) or showed a trend toward concordance (category III). Only 3%–4% of genes were discordant (category V). Additionally, the mean ratio on total RNA arrays of genes identified as outliers using amplified RNA but not total RNA was close to 2.0, suggesting that many of these genes represented real outliers that did not meet stringency criteria on total RNA arrays. Taken together, these data suggest that the genes identified as outliers using amplified RNA but not total RNA represent a greater sensitivity of amplified RNA to detect outliers using microarray analysis.

To validate this conclusion further, we analyzed the expression of outlying genes using real-time quantitative PCR. We considered gene expression in total RNA the "gold standard" for comparison. However, we also wished to evaluate the ability of real-time quantitative PCR to quantify relative gene expression in the amplified samples because total RNA is not always available from very small initial samples, such as those obtained by laser capture microdissection, and other techniques such as Northern blot analysis are hindered by hexamer-primed amplification. Genes from categories I–IV (not from category V) were validated by real-time quantitative PCR as being expressed differentially in all samples. It is unclear from our data whether the discordance seen in category V genes truly represents a bias of amplification. For example, the gene Prdx5 was identified as an outlier after one round of amplification (Figure 4F) but not after two rounds. Furthermore, real-time quantitative PCR suggested that microarray analysis overestimated the expression ratio for this gene using the single-round amplified material. Thus, the discordance seen among the data for this gene may be because of excessive interarray variation (rather than amplification bias). However, this phenomenon occurs for only 3%–4% of genes that had been identified as outliers using amplified RNA. Thus, the majority of outliers identified using amplified RNA appear to be valid, even if the same genes are not identified using total RNA. In addition, real-time quantitative PCR can be performed as a confirmatory technique on RNA samples after one or two rounds of amplification. Real-time quantitative PCR data that suggest that microarray analysis has overestimated the true expression ratio should raise suspicion that the gene may not be a true outlier. Thus, while previous studies (7,8,12) suggest the fidelity of overall expression profiles after amplification, our study is the first to address comprehensively the accuracy of identifying specific differentially expressed genes based on microarray analysis of amplified RNA.

In summary, our results indicate that one or two rounds of antisense amplification not only preserve the fidelity of total RNA-based microarray analysis but also actually improve the sensitivity for detecting outliers. This improvement is reflected in less interarray variability and in fewer genes that do not meet spot filtering criteria. This effect may be because of the improved quality of the RNA, which is transcribed in vitro under optimal conditions and is not subject to the effects of RNases and other factors present in cells and tissue specimens. Finally, we show not only that these findings are supported by the real-time quantitative PCR of total RNA but also that real-time quantitative PCR can be used with amplified RNA as a downstream validation technique. These findings suggest that up to two rounds of amplification can be used reliably when small initial amounts of total RNA are available. Furthermore, we advocate the general use of a single round of amplification before microarray analysis for all RNA specimens to allow the performance of multiple arrays without concerns about RNA quantity and to improve data quality and detection of differentially expressed genes.

REFERENCES

1. **Brown, P.O. and D. Botstein.** 1999. Exploring the new world of the genome with DNA microarrays. Nat. Genet. *21*:33-37.
2. **Collins, F.S.** 2001. Contemplating the end of the beginning. Genome Res. *11*:641-643.
3. **DeRisi, J., L. Penland, P.O. Brown, M.L. Bittner, P.S. Meltzer, M. Ray, Y. Chen, Y.A. Su, et al.** 1996. Use of a cDNA microarray to analyze gene expression patterns in human cancer. Nat. Genet. *14*:457-460.
4. **Florell, S.R., C.M. Coffin, J.A. Holden, J.W. Zimmermann, J.W. Gerwels, B.K. Summers, D.A. Jones, and S.A. Leachman.** 2001. Preservation of RNA for functional genomic studies: a multidisciplinary tumor bank protocol. Mod. Pathol. *14*:116-128.
5. **Gewolb, J.** 2001. Genomics. Animals line up to be sequenced. Science *293*:409-410.
6. **Kammula, U.S., K.H. Lee, A.I. Riker, E. Wang, G.A. Ohnmacht, S.A. Rosenberg and F.M. Marincola.** 1999. Functional analysis of antigen-specific T lymphocytes by serial measurement of gene expression in peripheral blood mononuclear cells and tumor specimens. J. Immunol. *163*:6867-6875.
7. **Luo, L., R.C. Salunga, H. Guo, A. Bittner, K.C. Joy, J.E. Galindo, H. Xiao, K.E. Rogers, et al.** 1999. Gene expression profiles of laser-captured adjacent neuronal subtypes. Nat. Med. *5*:117-122.
8. **Nagle, R.B.** 2001. New molecular approaches to tissue analysis. J. Histochem. Cytochem. *49*:1063-1064.
9. **Ohnmacht, G.A., E. Wang, S. Mocellin, A. Abati, A. Filie, P. Fetsch, A.I. Riker, U.S. Kammula, et al.** 2001. Short-term kinetics of tumor antigen expression in response to vaccination. J. Immunol. *167*:1809-1820.
10. **Ohyama, H., X. Zhang, Y. Kohno, I. Alevizos, M. Posner, D.T. Wong, and R. Todd.** 2000. Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. BioTechniques *29*:530-536.
11. **Phillips, J. and J.H. Eberwine.** 1996. Antisense RNA amplification: a linear amplification method for analyzing the mRNA population from single living cells. Methods *10*:283-288.
12. **Wang, E., L.D. Miller, G.A. Ohnmacht, E.T. Liu, and F.M. Marincola.** 2000. High-fidelity mRNA amplification for gene profiling. Nat. Biotechnol. *18*:457-459.
13. **Zogakis, T.G., N.G. Costouros, E.A. Kruger, S. Forbes, M. He, M. Qian, A.L. Feldman, W.D. Figg, et al.** Microarray gene expression profiling of angiogenesis inhibitors using the rat aortic ring assay. BioTechniques (In Press.)

Received 22 March 2002; accepted 17 May 2002.

Address correspondence to:

Dr. Steven K. Libutti
National Institutes of Health
Building 10, Room 3C428
Bethesda, MD 20892, USA
e-mail: slibutti@nih.gov