



A Highly Reproducible, Linear, and Automated Sample Preparation Method for DNA Microarrays

David R. Dorris, Ramesh Ramakrishnan, Dionisios Trakas, et al.

Genome Res. 2002 12: 976-984

Access the most recent version at doi:[10.1101/gr.227402](https://doi.org/10.1101/gr.227402)

References This article cites 23 articles, 9 of which can be accessed free at:
<http://genome.cshlp.org/content/12/6/976.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center is a white box with the text "LEARN MORE". On the right is a woman wearing a red and white superhero cape and mask, with the Cellecta logo (a green molecular structure) and the word "CELLECTA" below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Methods

A Highly Reproducible, Linear, and Automated Sample Preparation Method for DNA Microarrays

David R. Dorris,¹ Ramesh Ramakrishnan, Dionisios Trakas, Frank Dudzik, Richard Belval, Connie Zhao, Allen Nguyen, Marc Domanus, and Abhijit Mazumder

Motorola Life Sciences, Northbrook, Illinois 60062, USA

DNA microarrays are powerful tools to detect changes in transcript abundance in multiple samples in parallel. However, detection of differential transcript levels requires a reproducible sample (target) preparation method in addition to a high-performance microarray. Therefore, we optimized a target-preparation method that converts the poly(A)⁺ RNA fraction of total RNA into complementary DNA, then generates biotin-labeled complementary RNA from the cDNA. We measured the efficiency of incorporation of biotin-containing nucleotides by an enzymatic digestion, followed by resolution via analytical high-performance liquid chromatography (HPLC). When the target was hybridized to a sensitive and reproducible microarray platform, low coefficients of variation in both hybridization intensities and differential expression ratios across target preparations were observed. Nearly identical hybridization intensities and expression ratios are observed regardless of whether poly(A)⁺-enriched RNA or total RNA is used as the starting material. We show the ability to discern biological and production variability through the use of different lots of commercial samples as visualized by hierarchical clustering. Automation of the target-preparation procedure shows equivalence to the manual procedure, reproducible yields of target, and low variability as measured by hybridization to microarrays. Most importantly, RNA mixing experiments show a linear and quantitative amplification in probe hybridization signals for >6000 genes across the entire signal range.

DNA microarrays provide a powerful means to monitor the relative transcript abundance of many genes in parallel (Schena et al. 1995; Brown and Botstein 1999; Lockhart and Winzler 2000). Several factors are critical to obtain high-quality microarray data, including a reproducible, linear sample (target) preparation method, a sensitive and reproducible microarray platform, accurate image-acquisition software, and efficient, reliable data mining and computational tools. The ability to measure the variance in each step of the microarray assay process requires both high-quality manufacturing and a reproducible assay (Ramakrishnan et al. 2002). The ability to obtain high-quality data from a microarray platform enables an investigator to query the performance of the target-preparation procedure and its contribution to the overall variability because even low levels of variability in the target preparation can now be uncovered, measured, and tracked.

The available target-preparation methods can be divided into two groups: first-strand cDNA that is labeled or tagged with a capture sequence, or the generation of antisense RNA (aRNA) from double-stranded cDNA during an *in vitro* transcription (IVT) reaction. Labeled cDNA can be prepared via direct incorporation of a fluorophore-labeled nucleotide or through incorporation of an aminoallyl-labeled nucleotide, followed by coupling to a fluorophore containing an amine-reactive group to the aminoallyl nucleotide (Schena et al. 1995; for review, see Lockhart and Winzler 2000). Alternatively, the first-strand cDNA can be tagged with a capture

sequence that is used for subsequent detection steps (Stears et al. 2000). DNA microarrays containing short oligonucleotide probes (<35 nucleotides long) require more target for each hybridization, which requires an amplification method with smaller sample sizes. Typically, the generation of aRNA (aRNA is also commonly called complementary RNA or cRNA) is preceded by first-strand synthesis of cDNA using an oligonucleotide primer containing a bacteriophage T7 RNA polymerase promoter proximal to an oligo(dT) sequence (van Gelder et al. 1990; Eberwine et al. 1992; Lockhart et al. 1996). After second-strand cDNA synthesis and cDNA purification, an IVT reaction is performed using T7 RNA polymerase in the presence of labeled nucleotides. Alternatives to this labeling strategy produce unlabeled aRNA, followed by a cDNA synthesis in the presence of a fluorophore-labeled nucleotide (Wang et al. 2000). Any target preparation method requires a linear amplification of the available transcripts to be representative of the transcript population.

The starting material in microarray experiments can be either poly(A)⁺-enriched RNA or total RNA (Lockhart et al. 1996; Mahadevappa and Warrington 1999; Wang et al. 2000; Yue et al. 2001). Use of poly(A)⁺ RNA ensures a starting sample that is highly enriched in mRNA and possibly less contaminated by genomic DNA and total RNA. However, in cases where RNA quantities are limited, such as biopsy or laser-capture microdissection samples, it may not be possible to obtain sufficient quantities of total RNA to ensure that adequate quantities of poly(A)⁺ RNA can be purified from the total RNA. Furthermore, the additional handling and manipulation of the sample during the poly(A)⁺ RNA enrichment steps may introduce experimental artifacts. Recently, several reports have detailed methods and results starting with total

¹Corresponding author.

E-MAIL David.Dorris@email.mot.com; **FAX** (847) 714-7008.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.227402>.

RNA (Mahadevappa and Warrington 1999; Baugh et al. 2001; Zarrinkar et al. 2001). However, none of these reports has shown the correlation of either expression levels or expression ratios obtained when starting with poly(A)⁺ RNA compared with total RNA.

In this report, we have investigated the reproducibility of a target-preparation method that copies the messages into double-stranded cDNA, then amplifies the transcripts through the use of an IVT reaction (van Gelder et al. 1990; Eberwine et al. 1992; Lockhart et al. 1996). We present data wherein the coefficients of variation in the hybridization intensities are 6%–9% across target preparations from an automated target-preparation procedure. This low variability enables the demonstration of biological variability where it exists and of the low variability when data generated from poly(A)⁺-enriched RNA and total RNA are compared. Lastly, experiments in which total RNA is mixed from different tissues show that this target-amplification procedure is linear and authentically represents the starting material.

RESULTS

Quantitation of Biotinylated Nucleotides into aRNA

The generation of a high and reproducible specific activity of target aRNA is important for reproducible microarray data with maximal detection sensitivity. Therefore, the specific activity of the target sample was measured after optimization of the target-preparation procedure from that reported by others (Langer et al. 1981; van Gelder et al. 1990; Eberwine et al. 1992; Mahadevappa and Warrington 1999; Hill et al. 2000; Baugh et al. 2001). Published amplification protocols can produce a template-independent product in addition to amplifying the RNA of interest, thereby decreasing the specific activity of the final aRNA (Baugh et al. 2001). To ensure that no nonspecific target is produced using this method, multiple aRNA target preparations showed a size distribution similar to the cDNA preparation when analyzed by capillary electrophoresis (data not shown). The amount and uniformity of biotin incorporation across target preparations were measured by an analytical method comprised of an enzymatic digestion of the biotinylated, complex aRNA into mononucleosides, followed by resolution using HPLC. A typical chromatogram measuring absorbance at 260 nm shows excellent resolution of the four mononucleosides and the two biotin nucleoside peaks (Fig. 1). Biotinylated uridine and biotinylated cytosine were injected separately or coinjected with the digested aRNA to verify the identity of these peaks (data not shown). The incorporation frequency of the biotinylated cytosine and uridine can be calculated using the peak size as measured at 294 nm and the extinction coefficients of biotinylated uridine ($13,000 \text{ M}^{-1} \text{ cm}^{-1}$) and biotinylated cytosine ($9300 \text{ M}^{-1} \text{ cm}^{-1}$).

This calculation shows a fourfold better incorporation of biotinylated uridine than of the biotinylated cytosine. Furthermore, when combined with calculations for the unlabeled nucleosides, these calculations show that a biotinylated nucleotide was incorporated approximately every eight bases under these labeling conditions. Using this method, we have found high and reproducible specific activities across many tissues and target preparations (data not shown). Furthermore, these data showed the potential to reduce reagent costs in the assay through elimination of the biotinylated cytosine, owing to its lower incorporation rate, and through other measures. Based on experiments when biotinylated cytosine was eliminated from the in vitro transcription mix, biotinylated uridine was incorporated approximately every 10 bases, generating a high specific activity and comparable assay performance (data not shown). It is important to note that these data are all derived from digestion of a complex aRNA mixture and might not be representative of every transcript.

A Highly Reproducible Target-Preparation Method

Although a reproducible rate of incorporation of biotin-containing nucleotides across target preparations is desired, a reproducible transcriptional profile is the true goal. To test the reproducibility of the target-preparation procedure, four different aRNA preparations were generated from the same sample of placental total RNA (commercial source), then hybridized on one slide per preparation. Following the hybridization, detection, and image analysis, we calculated the coefficient of variation (CV) for the hybridization intensities for each probe (Fig. 2). The majority of the CVs in the hybridization intensities were below 30%, with an average CV (calculated using every single data point) of 16%. A total RNA preparation extracted from the human hepatocarcinoma cell line HepG2 by standard methods using chaotropic salts and phenol was used to generate two independent target preparations

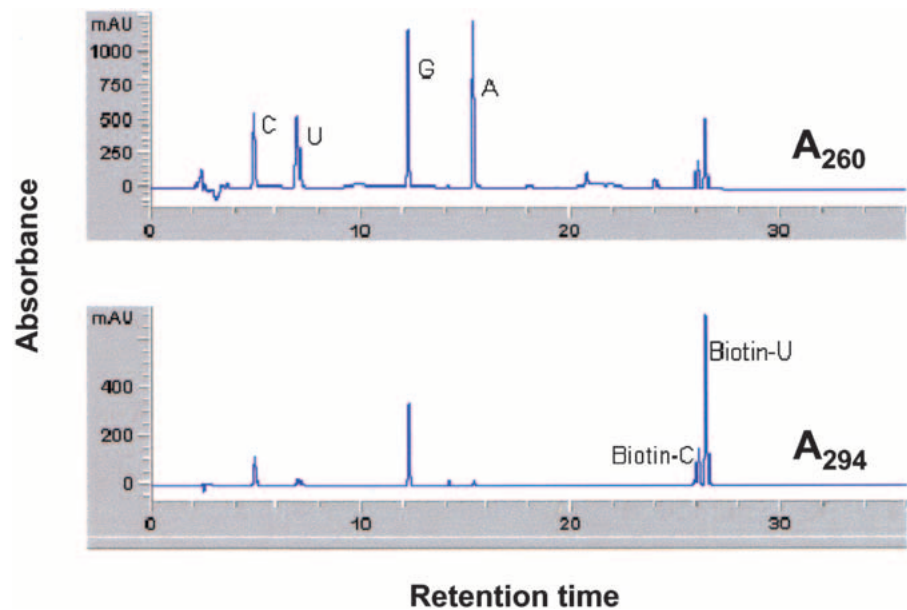


Figure 1 Biotin incorporation measurement. aRNA was digested to nucleosides, then separated by HPLC. The *upper* panel shows the absorbance profile of the heterocycles at 260 nm. The *lower* panel shows the absorbance profile at 294 nm. The biotin incorporation rate can be calculated using the area under each peak and the extinction coefficient of each nucleoside. The calculation indicates that a biotin-containing nucleotide is present approximately every 8 bases.

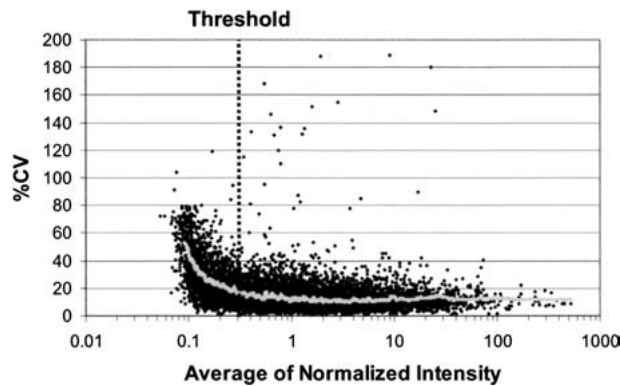


Figure 2 CV versus signal intensity graph. Four separate target preparations were hybridized to Human UniSet I slides and the coefficient of variation (CV) was plotted versus the mean signal intensity for each gene. The 100-point moving trend line shows that the average variability is low throughout the signal range. The average threshold, which represents the mean signal intensity plus 3 times the standard deviation for a set of bacterial control probes, is shown in red.

(see Methods). The expression ratio generated from these two aRNA preparations was plotted as a function of the signal intensity (ideally, every ratio should be equal to 1). Again, low variability throughout the entire signal range was observed (Fig. 3A). This variability is less than the variability reported in Hughes et al. (2001). Low variability in the target preparation should also be evident when differential expression ratios are examined. To address this issue, expression ratios were determined by comparing two separate aRNA preparations each for the HepG2 cell line and from liver tissue. A plot of the expression ratio changes in experiment 1 versus those observed in experiment 2 reveals a high correlation ($R = 0.984$) between the two experiments (Fig. 3B), as would be expected if there is low variability in the target-preparation methods.

A further demonstration of the low target-preparation variability was performed by generating 18 separate aRNA preps over three separate days from the same sample of total RNA across six different tissues or cell lines. The expression levels generated from each of these aRNA preps were then used to cluster the samples according to similarity using a hierarchical clustering algorithm (Fig. 4). We found that the samples clustered into five major clusters and every sample was clustered correctly based on its tissue or cell line of origin. Within the first cluster, there were two clusters corresponding to skeletal and cardiac muscle. Within the second cluster, there were two clusters corresponding to the two different Burkitt's lymphoma cell lines used (Namalwa and CA46V1). Within the liver cluster, there were three clusters corresponding to the three different livers tested from three individuals. The clustering also showed high levels of correlation between the replicate aRNA preps generated from the same RNA sample.

Automated Target Preparation

Reproducibility and the number of success rates of the target preparation could be further enhanced by automation of each reagent addition, incubations, and purification steps. A QIA-GEN robot with hardware and software modifications to perform all of the processes required for target preparation was used to generate aRNA (see Methods). The automated method

was found to be robust, with aRNA yields of 40–120 μg of aRNA generated from 5 μg of total RNA, and highly reproducible, with CVs in yields $\sim 10\%$ except for lymphoma tissue, which had a CV of 18% (Fig. 5A). Significantly, there were zero target-preparation failures (defined as a preparation with <30 μg of aRNA). This variability in yield between tissues could be due to the total RNA preparation or to the percentage of mRNA in each tissue's total RNA pool. Although these aRNA yields vary slightly less than a nonrobotic aRNA preparation technique obtained from multiple RNA sources, the

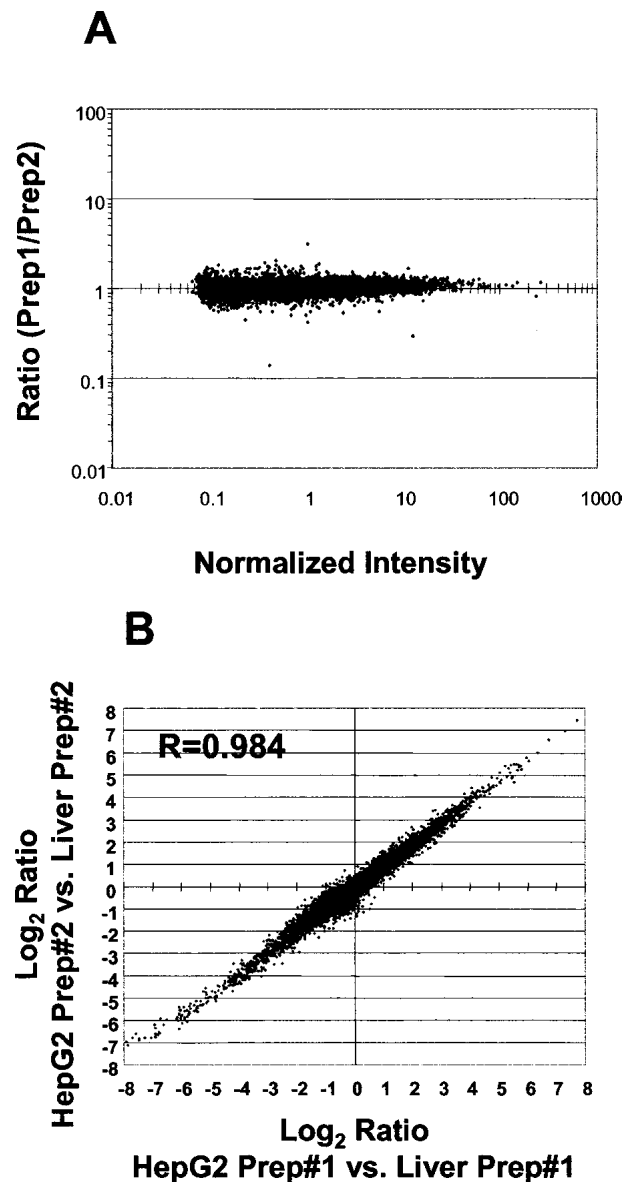


Figure 3 Target preparation variability. (A) aRNA target was prepared in two separate reactions from 5 μg of HepG2 total RNA, then hybridized to independent microarrays. The ratio of the signal intensity for each gene between preparations is plotted versus the signal intensity. (B) The signal intensities for each gene obtained from two aRNA target preparations in A were compared with two separate aRNA target preparations from the same liver total RNA sample (5 μg of input total RNA). The logarithm (base 2) ratio correlation would be equal to 1.0 if the target preparations and arrays were perfect.

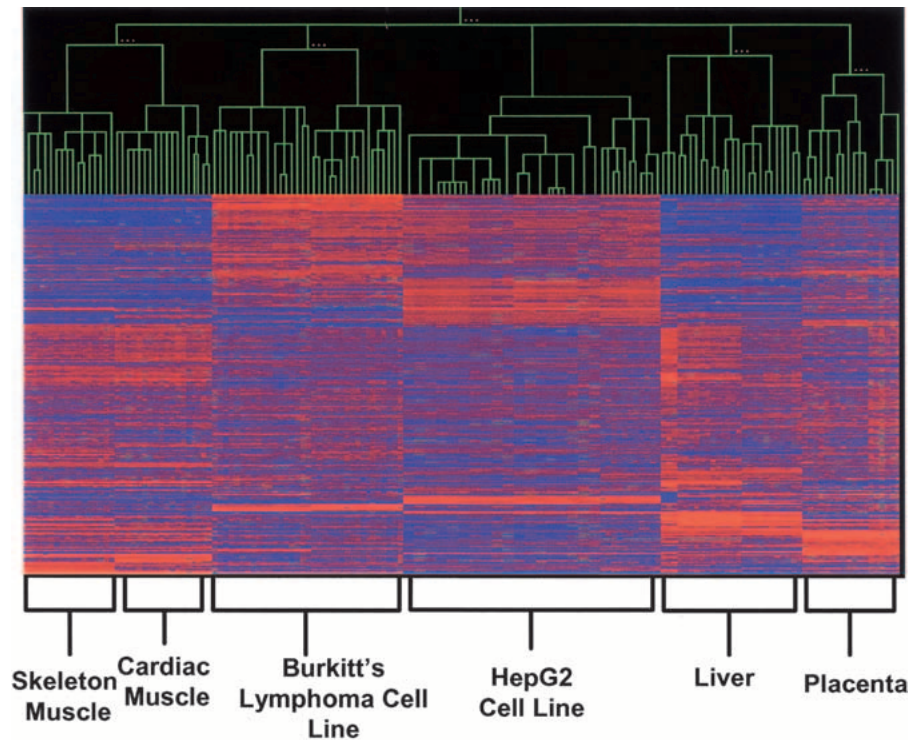


Figure 4 Hierarchical clustering. Microarray data from 161 Human UniSet I microarrays, corresponding to 84 aRNA target preparations from total RNA isolated from 6 different tissues (duplicate microarrays for 77 target preparations and single microarrays for 7 target preparations), were clustered using the Pearson correlation coefficient. The blue color indicates no expression, and red indicates a high expression level. There was a single source for both muscle types, two different Burkitt's lymphoma cell lines, two separate harvests of HepG2 cells, three liver sources, and two placental sources.

observation of variation is consistent with a published report (Zarrinkar et al. 2001).

The reproducibility of the transcriptional profiles was tested by selecting three random aRNA preparations from each tissue type for hybridization to microarrays. These three aRNA preparations per tissue were hybridized to duplicate microarrays for each target preparation, for a total of six microarrays per tissue. A microarray with sixfold probe redundancy for 1146 probes was used for the hybridizations to reduce the variability introduced by the microarray. For data analysis, the six spots corresponding to each probe were averaged to generate a single value for each probe per microarray to reduce variability attributable to the microarrays. A comparison of the variability versus the signal intensities for four of these tissues shows very low variability for all expression levels (Fig. 5B). In all nine tissues, the average CV for all 1146 probes is below 10% (data not shown).

We next determined whether the automated target preparations produced equivalent results to manual target preparations. If the methods produce different microarray results, an investigator would be precluded from comparing data generated with the two target-preparation methods. We generated 18 aRNA samples, 6 manually and 12 robotically, from the same sample of total RNA over 2 d. Each aRNA preparation was hybridized in duplicate, and Pearson correlation coefficients were calculated for each pairwise comparison (Table 1). From the data, it is apparent that there is an excellent linear correlation between hybridization intensities obtained when aRNA is generated from either the manual or

automated target-preparation method. A comparison of the ratio of the average intensities for each gene between the manual and robotic data shows the two methods to be equivalent (average ratio = 1.07 ± 0.06 ; maximum ratio = 1.36).

Total RNA versus Poly(A)⁺ RNA

Because the protocol developed in this study starts with total RNA instead of poly(A)⁺ RNA, we determined whether similar expression levels would be measured regardless of starting material. Poly(A)⁺ RNA was enriched from large pools of total RNA extracted from human liver or HepG2 cells by standard oligo(dT) chromatography. The total RNA and poly(A)⁺ RNA samples were used to generate aRNA, then hybridized to microarrays. A more detailed view of the data presented in Figure 4 shows that the signal intensities for each gene are very similar, regardless of whether the starting material was total or poly(A)⁺ RNA (Fig. 6A). A comparison of the differential expression ratios between the HepG2 and liver aRNA targets shows an excellent correlation between the total and poly(A)⁺ RNA ratios (Fig. 6B). Significantly,

the correlation coefficient is 0.95, and the slope of the regression line is 0.93, indicating very similar transcriptional profiles for aRNA target prepared from total RNA versus poly(A)⁺ RNA.

Sensitivity of Target-Preparation Procedure

The sensitivity of the amplification procedure was measured by adding exogenous spikes of poly(A)-tailed control mRNAs into the input RNA. The results show an equivalent sensitivity of the target-preparation procedure for poly(A)⁺ RNA versus total RNA when preparing a 1:150,000 or 1:6,000,000 mass:mass dilution, respectively (data not shown). This approximates the percentage of mRNA in total RNA as 2.5%. These comparisons indicate that the sensitivity of the amplification procedure is equivalent from total or poly(A)⁺ RNA as a function of the percentage of any individual mRNA in the mRNA pool.

A Linear Amplification Procedure

As is evident from these data, low target preparation variability should enable discrimination of samples where more subtle changes in transcript abundance are present. Thus, intensity changes should be reflective solely of the changes in transcript abundance if a linear relationship exists between the hybridization intensities and transcript levels, regardless of whether low or high expressors are examined. We confirmed this scenario by preparing RNA samples that were obtained by mixing brain and Burkitt's lymphoma total RNAs in

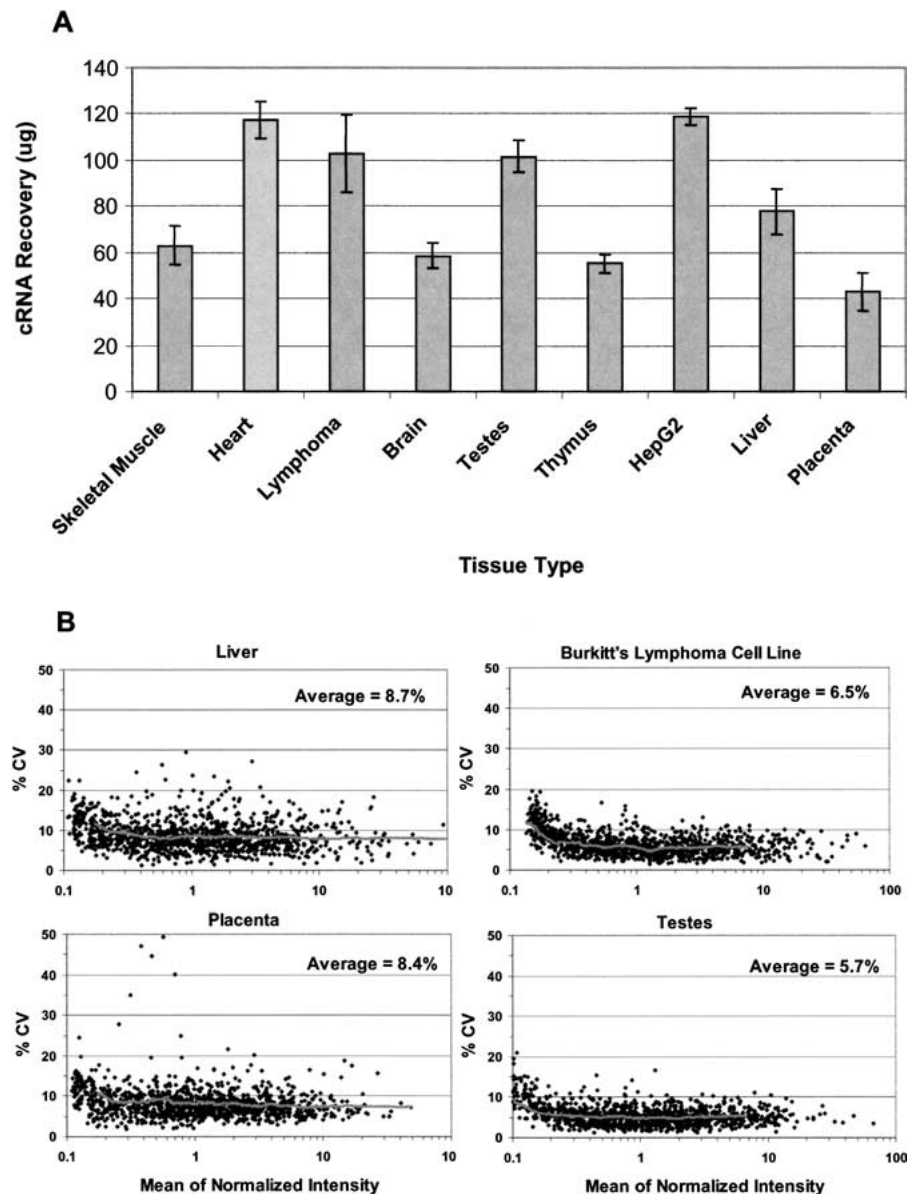


Figure 5 Robotic target preparation. (A) The average aRNA yields (\pm SD) from multiple tissues are shown (8 preparations/tissue). (B) The %CV is plotted versus the signal intensity for 3 separate target preparations as shown in A, each hybridized to a microarray containing 1100 genes spotted in six times each. The trend line represents the 100-point moving average of the %CV.

20:80, 40:60, 60:40, and 80:20 proportions. Each RNA mixture was used to generate aRNA and hybridized to oligonucleotide arrays. The results show that each gene shows a linear relationship for these RNA mixtures (Fig. 7A,B), although the slope is zero for nondifferentially expressed genes (data not shown). Further examination of the 3592 genes that are differentially expressed more than twofold in this mixture shows a linear response for each gene in the RNA mixtures (data not shown). The linearity of the intensities for each gene indicates that each transcript is amplified in a linear fashion, regardless of whether the gene was expressed at low, medium, or high levels in the two tissues. This same experimental approach was also used with mixtures of brain with placenta

and brain with lymphoma RNAs with equivalent linear results (data not shown).

An additional question raised is whether the target-amplification and microarray results show the correct fold change for these mixtures. To address this issue, we calculated the ratio of the intensities of individual genes in the 40:20, 80:40, and 80:20 brain:Burkitt's lymphoma mixtures for the 150 highest-expressed genes in brain that are absent in the Burkitt's lymphoma sample. These ratios indicate that, on average, the amplification is not only linear, but also faithfully represents the starting materials. The average amplification corresponds to the ratios of the input RNAs, and the standard deviation is very small (Fig. 7C). The presence of a very different, complex mixture of RNAs and subsequent cDNAs does not appear to affect the amplification procedure of individual probes.

DISCUSSION

Analysis of differential gene expression using DNA microarrays requires both accuracy and precision in every step of the manufacturing and assay process. We have presented data here to thoroughly analyze the performance of the target-preparation step of the assay process as monitored by both an analytical method and by microarray results. We provide evidence that a commonly used amplification scheme is linear for thousands of genes (Fig. 7). Additionally, we show that the slope of the linear amplification depends on the amount of mRNA and that the average amplification faithfully represents the starting material (Fig. 7C). In other words, a twofold increase in message corresponds to a twofold increase in signal. These data are in agreement with previous results

showing a linear correlation between in vitro prepared transcripts spiked into a starting poly(A)⁺ RNA sample at prescribed levels (Lockhart et al. 1996), although the data presented here measure the naturally occurring transcripts present in total RNA. Experiments in which we spike in vitro prepared mRNAs into total RNA have similar results to naturally occurring messages. Furthermore, we reproducibly detect spiked messages present at 83 parts per billion for multiple transcripts spiked into total RNA.

This target-preparation method is not only linear and sensitive, it is also very reproducible. The contribution of the manual target-preparation method to the overall variability is approximately one-third of the total variability, where the

Table 1. Correlation Coefficients Comparing Manual (M) and Robotic (R) Target-Preparation Methods

	M1	M1	M2	M2	M3	M3	M4	M4	M5	M5	M6	M6
R1	0.993	0.980	0.994	0.993	0.993	0.991	0.992	0.995	0.989	0.993	0.989	0.986
R1	0.985	0.967	0.983	0.979	0.979	0.977	0.977	0.984	0.975	0.977	0.974	0.971
R2	0.991	0.997	0.989	0.989	0.991	0.992	0.993	0.992	0.988	0.992	0.990	0.985
R2	0.987	0.976	0.987	0.989	0.992	0.993	0.994	0.991	0.990	0.995	0.991	0.987
R3	0.992	0.969	0.989	0.990	0.993	0.994	0.993	0.989	0.987	0.990	0.992	0.987
R3	0.992	0.972	0.990	0.991	0.993	0.993	0.993	0.990	0.987	0.991	0.991	0.987
R4	0.991	0.971	0.987	0.988	0.992	0.992	0.992	0.990	0.985	0.988	0.989	0.983
R4	0.991	0.968	0.988	0.989	0.992	0.992	0.992	0.988	0.985	0.987	0.990	0.985
R5	0.967	0.998	0.976	0.968	0.969	0.968	0.973	0.981	0.968	0.969	0.960	0.947
R5	0.985	0.987	0.988	0.987	0.990	0.990	0.993	0.993	0.986	0.989	0.985	0.978
R6	0.985	0.978	0.988	0.990	0.993	0.995	0.997	0.992	0.989	0.994	0.991	0.987
R6	0.982	0.975	0.984	0.988	0.991	0.994	0.995	0.990	0.988	0.993	0.990	0.986
R7	0.974	0.983	0.979	0.978	0.982	0.983	0.987	0.989	0.984	0.988	0.978	0.972
R7	0.981	0.981	0.986	0.985	0.987	0.986	0.989	0.992	0.985	0.989	0.982	0.977
R8	0.985	0.974	0.988	0.992	0.995	0.996	0.997	0.992	0.990	0.995	0.993	0.989
R8	0.989	0.975	0.991	0.993	0.996	0.997	0.998	0.993	0.990	0.995	0.994	0.990
R9	0.979	0.966	0.982	0.987	0.990	0.992	0.993	0.985	0.987	0.991	0.989	0.987
R9	0.980	0.974	0.981	0.985	0.989	0.992	0.994	0.988	0.986	0.991	0.987	0.983
R10	0.983	0.979	0.986	0.987	0.991	0.992	0.994	0.991	0.988	0.992	0.988	0.983
R10	0.985	0.986	0.983	0.986	0.991	0.993	0.993	0.988	0.987	0.991	0.990	0.988
R11	0.989	0.976	0.993	0.994	0.993	0.993	0.994	0.992	0.987	0.990	0.991	0.986
R11	0.987	0.972	0.993	0.994	0.993	0.992	0.993	0.991	0.986	0.990	0.990	0.987
R12	0.979	0.967	0.983	0.985	0.983	0.982	0.984	0.982	0.979	0.982	0.980	0.977
R12	0.985	0.975	0.988	0.991	0.992	0.993	0.995	0.991	0.988	0.992	0.990	0.985

additional variability is introduced by the slide, hybridization, scanning, and spot quantitation procedures, as calculated using statistical blocking by ANOVA (data not shown). This conclusion differs from a published report that concluded that the target preparation had no contribution to the total variance (Yue et al. 2001). This discrepancy is most likely due to differences in the statistical techniques used to identify the sources of variability. Analysis of the robotic target-preparation method indicates a lower amount of variability (Fig. 5B). Part of the target-preparation variance occurs during the preparation of the reagent mixtures used for the different enzymatic steps (data not shown). Additional gains in reproducibility will most likely occur with reduced handling by the end user through the use of dried reagents and premade reagent mixes.

The robotic target-preparation procedure greatly increases the throughput of the entire microarray assay. Although the actual amplification technique described here is optimized to reduce the labor requirements for sample preparation, the throughput for the manual process is maximally 24 preparations/d per person. The robotic preparation method increases that throughput to 192 preparations/d, with less than a 1% failure rate (<30 μ g of aRNA produced) when starting with 5 μ g of total RNA input (D. Dorris and F. Dudzik, unpubl.). The throughput and lower variability of the robotic method are a useful match for constructing expression profiling databases, for large experiments, and for molecular classification of cell and tissue types (DeRisi et al. 1997; Marton et al. 1998; Golub et al. 1999; Iyer et al. 1999; Alizadeh et al. 2000; Bittner et al. 2000; Hughes et al. 2000; Perou et al. 2000; Shoemaker et al. 2001). Identifying molecular markers for tissue types or disease states requires a stable target-preparation method as part of an assay that provides sensitivity and reproducibility, combined with a high-quality microarray (Ramakrishnan et al. 2002).

The amount, quality, and type of the starting material are of paramount importance. Isolation of large quantities of

RNA from tissue culture experiments or large tissue samples permits further purification of a sample to poly(A)⁺ RNA. Unfortunately, this type of experiment will likely be in the minority in the future as expression profiling moves to biopsy samples, cell-sorted samples, and eventually laser-capture microdissection samples. Some microarray experiments use poly(A)⁺ RNA instead of total RNA because of background and genomic DNA contamination issues (Stears et al. 2000). Alternatively, some methods suggest using >25 μ g of total RNA as the starting material for first-strand cDNA-labeling methods (Yue et al. 2001). We can routinely start with <1 μ g of total RNA input for the amplification procedure, followed by a single hybridization (D. Dorris and D. Trakas, unpubl.). The data presented in Figure 6 indicate that using total RNA as the starting material produces equivalent results to poly(A)⁺ RNA. Therefore, using the linear amplification procedure described here when starting with total RNA is appropriate to represent the transcriptional profile of the cells being tested. This detailed comparison of the transcriptional profiles of total RNA versus poly(A)⁺ RNA indicates that the >95% of the content in total RNA that is represented by transfer and ribosomal RNAs does not have a deleterious effect on the amplification of the mRNA when using a modification of the original approach of Eberwine and colleagues (van Gelder et al. 1990; Eberwine et al. 1992). Whether there is an effect of tRNA and rRNA on random priming approaches of first-strand cDNA from total RNA labeling methods is unclear, yet possibly important to fully define the variability introduced when preparing samples for DNA microarrays using this method.

METHODS

RNA Preparation

Total RNA was purchased for all tissues and the Burkitt's lymphoma cell lines (Ambion; BioChain; Clontech), or prepared from the hepatocarcinoma cell line HepG2 using the TriZOL method (Invitrogen). The standard TriZOL RNA extraction

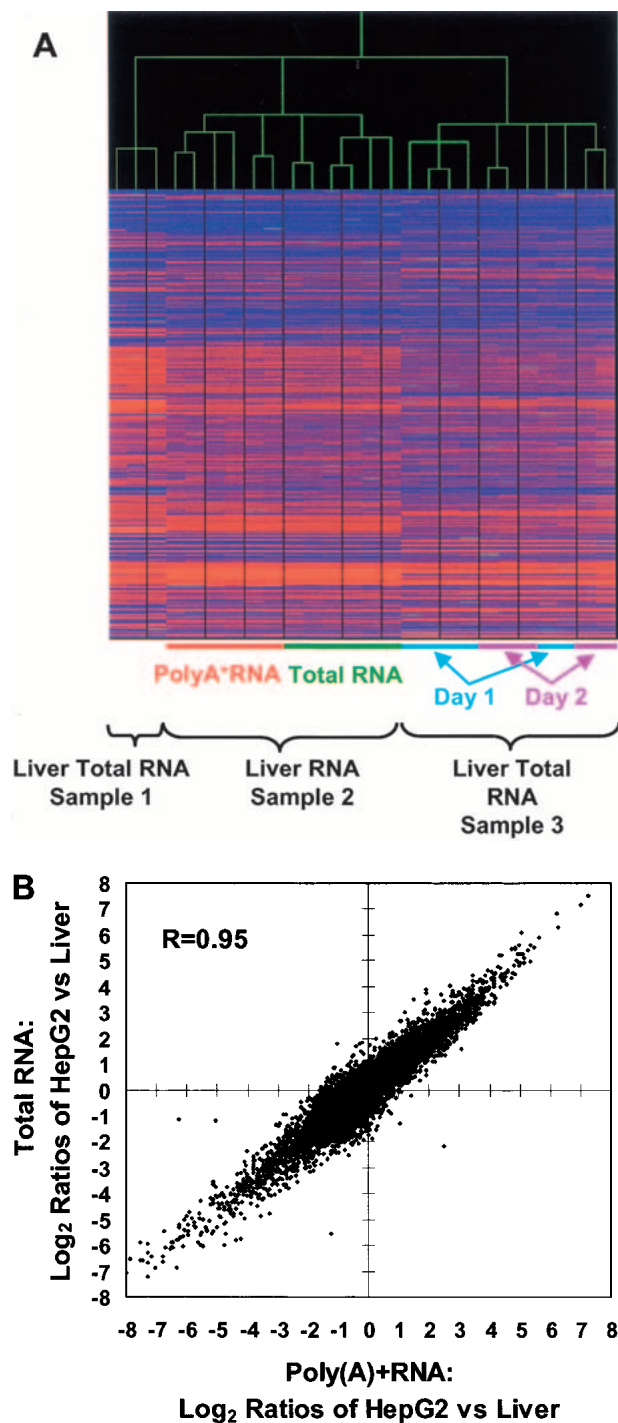


Figure 6 Total RNA versus poly(A)⁺ RNA. Poly(A)⁺ RNA was enriched from a portion of a large pool of total RNA isolated from a single liver sample or from HepG2 cells. Multiple aRNA preparations were generated from both the poly(A)⁺ RNA and total RNA samples and hybridized to Human UniSet I microarrays in duplicate. (A) The microarray results are included in the hierarchical clustering diagram shown in Figure 4, which is shown in more detail here to highlight the comparison of aRNA prepared from poly(A)⁺ RNA. (B) The logarithm (base 2) of the differential expression ratios of HepG2 versus liver are plotted for total RNA compared with poly(A)⁺ RNA. Every ratio measured is plotted from a single microarray per condition (4 aRNA preparations and 4 microarrays total). The correlation coefficient (*R*) was calculated to be 0.95.

protocol was modified by adding a chloroform extraction step following the TriZOL extraction step. HepG2 cells were grown in MEM α plus 10% fetal calf serum in the presence of penicillin and streptomycin. Poly(A)⁺ RNA was prepared using a MEGApure kit according to the manufacturer's instructions (Ambion).

Manual Target Preparation

We added 5 μ g of total RNA to a reaction mix in a final volume of 12 μ L containing 0.5 pmole/ μ L T7-(dT)₂₄ oligonucleotide primer. The mixture was incubated at 70°C for 10 min, then chilled on ice. With the mixture remaining on ice, 4 μ L of 5 \times first-strand buffer, 2 μ L of 0.1 M DTT, 1 μ L of 10 mM dNTP mix, and 1 μ L of Superscript II RNase H⁻ reverse transcriptase (200 U/ μ L) were added to a final volume of 20 μ L, and the mixture was incubated in a 42°C waterbath for 1 h. Second-strand cDNA was synthesized in a final volume of 150 μ L, in a mixture containing 30 μ L of 5 \times second-strand buffer, 3 μ L of 10 mM dNTP mix, 4 μ L of *Escherichia coli* DNA polymerase I (10 U/ μ L), and 1 μ L of RNase H (2 U/ μ L) at 16°C for 2 h. The cDNA was purified using a QIAGEN QIAquick purification kit, dried down by lyophilization, and resuspended in an IVT reaction mix containing 3.0 μ L of nuclease-free water, 4.0 μ L of 10 \times reaction buffer, 4.0 μ L of 75 mM ATP, 4.0 μ L of 75 mM GTP, 3.0 μ L of 75 mM CTP, 3.0 μ L of 75 mM UTP, 7.5 μ L of 10 mM Biotin 11-CTP, 7.5 μ L of 10 mM Biotin 11-UTP, and 4.0 μ L of enzyme mix. The reaction mix was incubated at 37°C for 14 h, and the aRNA target was purified using an RNeasy Kit (QIAGEN). aRNA yield was quantitated by measuring the UV absorbance at 260 nm. The aRNA was fragmented in 40 mM Tris-acetate (TrisOAc; pH 7.9), 100 mM KOAc, and 31.5 mM MgOAc, at 94°C for 20 min immediately before hybridization with no further purification steps. This typically resulted in fragmented target with a size range between 100 and 200 bp. Target preparations containing spiked transcripts with exogenously added spikes (prepared in vitro) were added at a mass:mass ratio in a range from 1:6,000,000 to 1:200,000 into total RNA, or a range from 1:300,000 to 1:1000 into poly(A)⁺ RNA.

Automated Target Preparation

The automated target preparations followed the above protocol with the following modifications. All incubations were performed in microtiter plates in a thermal cycler containing a heated lid. All liquid and enzyme solution additions and purifications were performed on a QIAGEN BioRobot (model 9604 or 3000). cDNA purification and aRNA purification were performed using QIAquick 96 PCR purification kits and RNeasy96 kits (QIAGEN), respectively.

Digestion and Chromatography of aRNA

To generate nucleotide monophosphates, 4 U of P1 nuclease were used to digest 20–50 μ g of aRNA. The enzyme was incubated with the aRNA at 55°C for 6 h, then at 37°C for 6 h in the presence of 10 U of calf intestine alkaline phosphatase to generate the nucleosides. The digested products were purified using Microcon YM-3 columns followed by centrifugation at 8000g for 30–60 min. The mix was then concentrated using a SpeedVac to 100 μ L. This solution was analyzed on an HPLC column equilibrated with 0.03 M TEAA (Solvent A) at a flow rate of 1 mL/min. The following gradient was used: 0%–1% Solvent B (95% AcCN, 5% Solvent A1) over 5 min, 1%–15% Solvent B over the next 15 min, 15%–45% Solvent B over the next 30 min, 45%–100% Solvent B over the next 20 min, and hold at 100% Solvent B for 2 min. The concentration of the heterocycles was determined by the absorbance values at 260 nm (the wavelength where maximal absorption occurs for the heterocycles), and the biotin-containing nucleoside concentrations were determined by the absorbance values at 294 nm

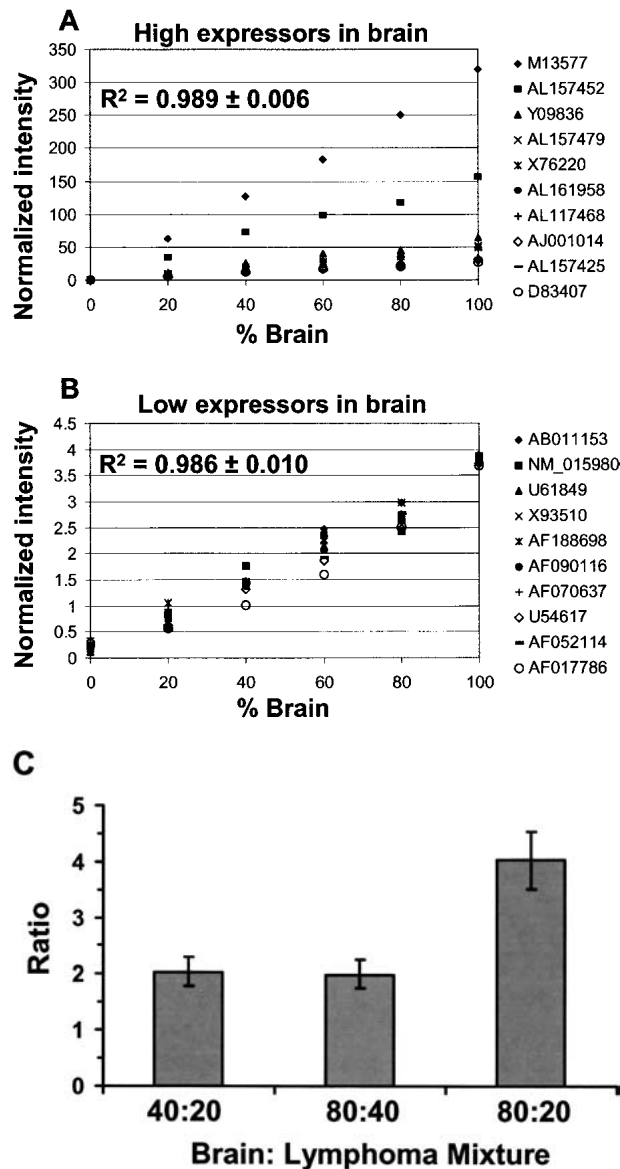


Figure 7 Linearity of amplification. Total RNA from brain and a Burkitt's lymphoma cell line (CA46V1) was mixed at 80:20, 60:40, 40:60, and 20:80 mass:mass ratios, then used to generate aRNA. These aRNA preparations were hybridized in triplicate to Human UniSet I arrays. (A) The average signal intensities for 10 genes expressed at high levels in brain tissue, but not present in lymphoma tissue, are plotted versus the percentage of brain total RNA in the original mixture. The average R^2 value (\pm SD) for each set of the 10 genes plotted is shown for A and B. (B) The average signal intensities for 10 genes expressed at low levels in brain tissue, but not present in lymphoma tissue, are plotted versus the percentage of brain total RNA in the original mixture. (C) The average brain:lymphoma ratio of the signal from the indicated mixture for the 150 highest expressed genes in brain that are not expressed in the Burkitt's lymphoma cell line. The error bars represent the standard deviation.

(the wavelength where maximal absorption occurs for biotinylated cytosine and biotinylated uridine).

Array Hybridization and Processing

We used 10 μ g of fragmented target aRNA for hybridization of

each UniSet Human I Expression Bioarray chip containing 9589 probes (these 9589 probes represent 9203 unique accession numbers [genes], corresponding to ~8935 unique clusters and 386 control probes, selected initially from GenBank Uni-gene build #125) or for hybridization to a microarray containing probes corresponding to 1100 human genes, each spotted 6 times per array (Motorola Life Sciences). All probes on these microarrays are 30-mer oligonucleotides spotted by piezoelectric technologies and covalently attached to a polymeric surface (Ramakrishnan et al. 2002). These microarrays were hybridized, washed, and processed using a direct detection method of the biotin-containing transcripts by a Streptavidin-Alexa647 conjugate as described (Ramakrishnan et al. 2002). Processed slides were scanned using an Axon GenePix Scanner with the laser set to 635 nm, a PMT voltage of 600, and a scan resolution of 10 μ m.

Data Analysis

Slides were scanned using CodeLink Scanning Software (Motorola Life Sciences), and images for each slide were quantitated using the CodeLink Expression Analysis Software (Motorola Life Sciences). Signal intensities for each spot were calculated by summation of the pixel intensities for each spot, then the local background (based on the median pixel intensity of the area surrounding each spot) was subtracted. Whole array data normalization was performed independently for each slide by dividing each spot's intensity (after background subtraction) by the median signal intensity of all test probes. Hierarchical clustering was performed using the Pearson correlation coefficient in GeneSpring software (Silicon Genetics).

ACKNOWLEDGMENTS

We thank Geetha Rajavalu for help with ANOVA blocking.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. 2000. Distinct types of B-cell lymphoma identified by gene expression profiling. *Nature* **403**: 503–511.
- Baugh, L.R., Hill, A.A., Brown, E.L., and Hunter C. P. 2001. Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* **29**: E29.
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., et al. 2000. Molecular classification of cutaneous malignant melanoma by gene expression profiling. *Nature* **406**: 536–540.
- Brown, P.O. and Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* **21** (Suppl. 1): 33–37.
- DeRisi, J.L., Iyer, V.R., and Brown, P.O. 1997. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680–686.
- Eberwine, J., Yeh, H., Miyashiro, K., Cao, Y., Nair, S., Finnell, R., Kettel, M., and Coleman, P. 1992. Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci.* **89**: 3010–3014.
- Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286**: 531–537.
- Hill, A.A., Hunter, C.P., Tsung, B.T., Tucker-Kellogg, G., and Brown, E.L. 2000. Genomic analysis of gene expression in *C. elegans*. *Science* **290**: 809–812.
- Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. 2000. Functional discovery via a compendium of expression profiles. *Cell* **102**: 109–126.
- Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J.,

- Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotech.* **19**: 342–347.
- Iyer, V.R., Eisen, M.B., Ross, D.T., Schuler, G., Moore, T., Lee, J.C.F., Trent, J.M., Staudt, L.M., Hudson, Jr., J., Boguski, M.S., et al. 1999. The transcriptional program in the response of human fibroblasts to serum. *Science* **283**: 83–87.
- Langer, P.R., Waldrop, A.A., and Ward, D.C. 1981. Enzymatic synthesis of biotin-labeled polynucleotides: Novel nucleic acid affinity probes. *Proc. Natl. Acad. Sci.* **78**: 6633–6637.
- Lockhart, D.J. and Winzler, E.A. 2000. Genomics, gene expression, and DNA arrays. *Nature* **405**: 827–836.
- Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotech.* **14**: 1675–1680.
- Mahadevappa, M. and Warrington, J.A. 1999. A high-density probe array sample preparation method using 10- to 100-fold fewer cells. *Nat. Biotech.* **17**: 1134–1136.
- Marton, M.J., DeRisi, J.L., Bennett, H.A., Iyer, V.R., Meyer, M.R., Roberts, C., Stoughton, R., Burchard, J., Slade, D., Dai, H., et al. 1998. Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat. Med.* **4**: 1293–1301.
- Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. 2000. Molecular portraits of human breast tumors. *Nature* **406**: 747–752.
- Ramakrishnan, R., Dorris, D., Lublinsky, A., Nguyen, A., Domanus, M., Prokhorova, A., Gieser, L., Touma, E., Lockner, R., Tata, M., et al. 2002. Development and use of analytical tools in the dissection and optimization of microarray performance. *Nucleic Acids Res.* **30**: e30.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Shoemaker, D.D., Schadt, E.E., Armour, C.D., He, Y.D., Garrett-Engle, P., McDonagh, P.D., Loerch, P.M., Leonardson, A., Lum, P.Y., Cavet, G., et al. 2001. Experimental annotation of the human genome using microarray technology. *Nature* **409**: 922–927.
- Stears, R.L., Getts, R.C., and Gullans, S.R. 2000. A novel, sensitive detection system for high-density microarrays using dendrimer technology. *Physiol. Genomics* **3**: 93–99.
- van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci.* **87**: 1663–1667.
- Wang, E., Miller, L.D., Ohnmacht, G.A., Liu, E.T., and Marincola, F.M. 2000. High fidelity mRNA amplification for gene profiling. *Nat. Biotech.* **18**: 457–459.
- Yue, H., Eastman, P.S., Wang, B.B., Minor, J., Doctolero, M.H., Nuttall, R.L., Stack, R., Becker, J.W., Montgomery, J.R., Vainer, M., et al. 2001. An evaluation of the performance of cDNA microarrays for detecting changes in global gene expression. *Nucleic Acids Res.* **29**: E41.
- Zarrinkar, P.P., Mainquist, J.K., Zamora, M., Stern, D., Welsh, J.B., Sapinoso, L.M., Hampton, G.M., and Lockhart, D.J. 2001. Arrays of arrays for high-throughput gene expression profiling. *Genome Res.* **11**: 1256–1261.

Received December 13, 2001; accepted in revised form April 11, 2002.