

Short Technical Report

Toward a Universal Standard: Comparing Two Methods for Standardizing Spotted Microarray Data

BioTechniques 32:1310-1314 (June 2002)

Michael Ryan Weil, Timothy Macatee, and Harold R. Garner

University of Texas Southwestern Medical Center, Dallas, TX, USA

ABSTRACT

DNA microarray technology has allowed the transcriptome to be studied to a depth that was inconceivable only 10 years ago. Until recently these studies were isolated because, without a universal standard, the results from experiment to experiment and laboratory to laboratory were not directly comparable. For human microarrays, this problem has been addressed by numerous methods, but only two are truly universal. The first method uses genomic DNA as a standard for comparison since it is, by definition, complete and universally available. The second method employs a highly representative total RNA pool such as the one currently available from Stratagene. To determine the advantages and disadvantages of both methods, they were directly compared by hybridization to the University of Texas Southwestern Medical Center's 4000- or 10 800-member human cDNA array, using typical microarray techniques. The labeled analytes were 2 µg normal human genomic DNA labeled by nick translation or 20 µg total RNA pool labeled by reverse transcription. The resulting data were then background-subtracted, analyzed, and the number of spots above a background threshold

was compared in each sample. Using the McNemar test and a Yate's correction with one degree of freedom, the samples were statistically identical with $X^2 = 3.72$.

INTRODUCTION

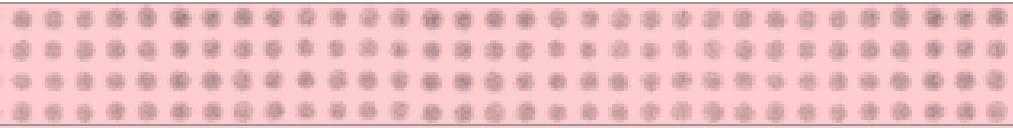
The study of the transcriptome has matured in the last five years from a slow-moving field based on traditional molecular biology to one of the most progressive fields in biology. This was in part made possible by the invention of the various types of DNA microarray technologies that allow for the elucidation of the comparative expression states of thousands of genes (13). It is not the intention of this report to cover the scope of DNA microarray techniques and technologies because several excellent reviews on the subject are already available (5), but a brief introduction is necessary.

There are many types of DNA microarray, but the two major platforms are the GeneChip[®] oligonucleotide microarray (Affymetrix, Santa Clara, CA, USA), in which a known oligonucleotide (probe) is synthesized directly on glass arrays, and DNA-spotted microarrays, in which the probe, generally a PCR product, is spotted onto a substrate by precision robotics (6,9). In both cases, an experiment consists of hybridization of a fluorescently labeled cDNA (analyte, target, or uncharacterized sample) against the probe. The raw data are

obtained by measuring the resulting intensity values of the spots by fluorescent microscopy with laser excitation.

Oligonucleotide microarrays use a single-color labeled analyte to measure the relative abundance of transcripts in a cDNA sample. Through manufacturing reproducibility and the use of internal controls, the analyte is able to compute relative expression changes by comparing one array to another. DNA arrays have the advantage of being able to hybridize to several differentially labeled target samples, thus allowing the comparison of relative expression changes on a single array; however, the comparison of expression levels from slide to slide is problematic.

Array-to-array comparison presents a challenge because even different preparations of the same sample can show large changes in gene expression (7). In the case of oligonucleotide microarrays, the quality of the chip and the large number of internal standards are sufficient to compare experiments using array-specific software (4,16). However, for spotted arrays where the quality of each spot is unknown and varies across the array, an additional and more rigorous standardization process is needed to supplement the internal standards. This standardization takes the form of a second analyte or universal reference standard that is used in all of the experiments. This universal standard allows for the comparison of experimental samples by



providing a common set of spots for normalization and a means for internal quality control. To compare a universal reference sample for standardization between arrays, a ratio of the ratios is taken so that the value of the standard cancels out, and the experiments can be compared. Unfortunately, if the intensity value of the spots in the reference sample is not significantly above the array's background or threshold, then the spot and the transcript it represents cannot be accurately compared across experiments.

To overcome this, many analysis schemes have been developed to set a user-defined lower boundary based on the mean intensity, plus 1–2 standard deviations of some set of negative control spots, generally buffer blanks or heterologous DNA. This is still not a direct comparison because the experimental variation is not well accounted for and is often unreliable (1,17). The lack of expression in the reference channel was a problem in early array experiments because the reference sample used was usually a control cDNA sample. Therefore, changes in the relative expression of genes that were only expressed in a single sample or color could not be reliably measured. As early as 1998, this need to compare experiments forced various laboratories to look for better reference standards, and the search continues today (3,8).

With numerous array-based transcriptome surveying projects already underway, most notably the National Institutes of Health Microarray Project (μ AP), finding a universal standard for comparing the results has become paramount (12). For human microarrays, this problem has been addressed by numerous methods such as using a common primer to determine how much DNA is present in each spot, using a mixture of the spotted clones to determine the baseline, or simply "spiking" the reference sample with RNAs of interest (2). However, these methods are inadequate because they are array-specific or do not test the labeling and hybridization characteristics directly. A good universal standard must hybridize competitively to the same sites as the experimental sample, thereby providing a direct measure of the hybridization process for all of the genes spotted

on the array (3). In addition, it must be widely available and usable on any expression microarray. Because changes in gene expression are seen in every RNA preparation, any RNA standard should be abundant enough to last through a complete experimental run, and beyond if possible.

In a seminal review, Eisen and Brown (2) not only stated the need for a universal reference but also named two methods to accomplish this. The first method uses genomic DNA because it is complete and universally available. The second employs a highly representative RNA pool such as that currently available from Stratagene (La Jolla, CA, USA). Both standardization methods should hybridize to all of the DNAs on an array at a level well above background and be highly reproducible. Here we summarize an experiment to compare the two methods and determine if one is superior.

MATERIALS AND METHODS

Genomic DNA Labeling

High molecular weight genomic DNA was extracted from normal human lung fibroblasts (GM 01604) or was purchased from Invitrogen (Carlsbad, CA, USA). The labeling was done using the Nick Translation kit (Promega, Madison, WI, USA) with 2 μ g genomic DNA. This DNA concentration was experimentally determined (concentrations from 0.5 to 5 μ g were tested) to minimize DNA-based suppression of the RNA signal while maximizing the number of spots above threshold (data not shown). The protocol supplied with the kit was used with two exceptions. First, no unlabeled dCTP was used in the 10 mmol dNTP mixture; instead, 1 μ L of the appropriate cyanine (Cy) dye-conjugated dCTP (Amersham Biosciences, Piscataway, NJ, USA) was added to each 50 μ L reaction (14,15). Second, the reaction was allowed to run for 7.5 h, instead of the specified 15 min, to increase the incorporation of dye and reduce the average fragment size.

RNA Labeling

The total RNA pool was purchased from Stratagene. The pool consists of

10 proprietary cell lines representative of most of the body's tissues and therefore representative of most of the expressed sequences (10). Twenty micrograms of the total RNA pool were reverse-transcribed using SUPERSCRIPT® II and Oligo (dT) 12–18 from Invitrogen. The labeling reaction was performed using a standard protocol (2). The variations from this protocol were a reduction in the reaction volume to 40 μ L and a 1-h primer annealing step before the addition of reverse transcriptase to increase the specificity.

Array Production

The arrays used in this report were produced in the University of Texas Southwestern (UTSW) Array Core using the PCR products of sequence-verified clones (Research Genetics, Huntsville, AL, USA) and the universal vector primers (forward, 5'-CTGCAAGGCG ATTAAGTTGGGTAAC-3', and reverse, 5'-GTGAGCGGATAACAATTTTCAC-ACAGGAAACAGC-3'). A complete clone list is available at http://lethargy.swmed.edu/pre_computed.asp. After the size and purity (single band) were verified using gel electrophoresis, the PCR products were resuspended in 7% DMSO and printed at a pitch of 0.28 mm on poly-D-lysine-coated slides using MAGNA™, a custom-built spotting robot available from Bioautomation (Plano, TX, USA). The detailed protocols, array analysis template files, and the materials list are available at <http://microarray.swmed.edu>.

Array Clean Up and Hybridization

The labeling reactions were purified using 30K Microcon® spin columns (Millipore, Bedford, MA, USA) with two 400- μ L TE rinses. The elution volume was less than or equal to 16 μ L, with any missing volume replaced with TE to bring the final combined volume to 32 μ L. The hybridization buffer consisted of 8 μ L 20 \times SSC and 2 μ L 10% SDS passed through a 0.22- μ m filter, to which 3.2 μ L 10 mg/mL yeast tRNA were added for a total volume of 45.2 μ L. After boiling for 2 min, the solution was pipetted onto one of the UTSW arrays, either a 10800- or 4000-member array, and covered with a 24 \times 60 mm

coverslip. The array was placed in a sealed hybridization chamber (Telechem, Sunnyvale, CA, USA), with 10 μ L 3 \times SSC in the wells to maintain the humidity. The chamber was then sealed and placed in a 61 °C water bath for 16 h.

Washing and Scanning

After the hybridization chamber was removed from the water bath, the chamber was immediately opened and the array was placed in the first wash buffer (2 \times SSC, 0.1% SDS.) The cover slide was floated off in the first wash buffer, and then the container was gently agitated for 5 min. The array was washed in the next two wash buffers (0.4 \times and 0.2 \times SSC, respectively) as before. The array was dried by centrifugation at 155 \times g for 3 min at room temperature and scanned on a GenePix™ 4000b two-color scanner (Axon Instruments, Union City, CA, USA) at 10 μ m resolution. The pixel-intensity extraction and spot flagging were done with the GenePix 3.0 software package (Axon Instruments), and the output was converted to a tab-delimited text file.

Analysis

The data file was imported into a custom Microsoft® Excel® worksheet developed at UTSW. The values of the mean signal intensity from each spot were subtracted from the local mean background intensity. Spots that were not above the threshold (calculated by the mean of the blanks plus one standard deviation of the blanks) were filtered out. The data was then analyzed for the number of spots above this threshold, reproducibility, and the threshold value. Using the DAG STAT statistical package for Excel (The Mental Health Research Institute, Victoria, Australia), we applied the McNemar test with a Yate's correction to find the concordance between the two reference standards (11).

RESULTS AND DISCUSSION

Nine arrays (seven of the 4000-member array and two of the 10800-member array) were hybridized to compare the two standards, with very similar results. Here we will focus our

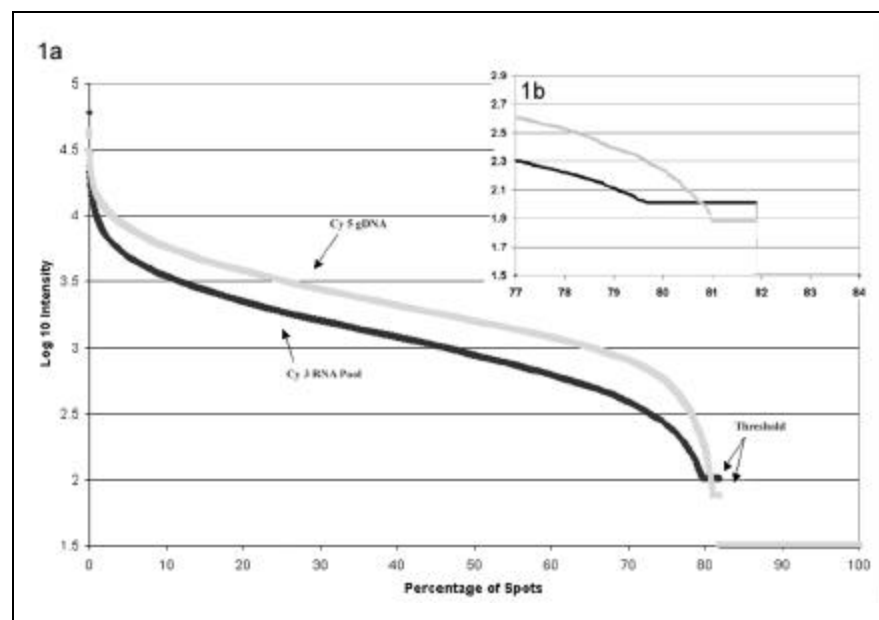
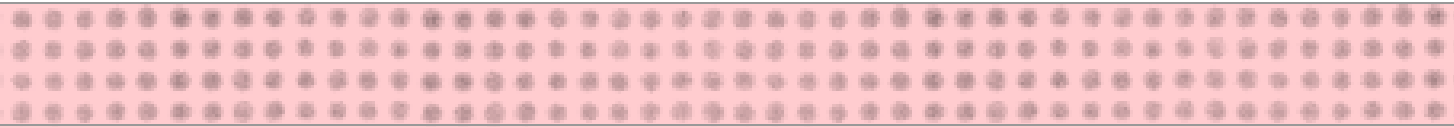


Figure 1. The log₁₀ of the background-subtracted spot intensities sorted by value shows the minimal difference between the two reference standards. One hundred percent corresponds to 10800 spots on the microarray. (A) The gray line represents the genomic DNA labeled in Cy5, and the black line represents the RNA pool labeled in Cy3. (B) A blowup of the low-intensity threshold region of the curve better illustrates the small difference between the two standards. The flattening of the curve and the length of the tails are measures of the number of spots above the threshold and the reproducibility of the standards, respectively. The line at the bottom at 1.5 log₁₀ intensity is where neither standard is above threshold in either channel and therefore is a measure of the coverage failure.



analysis on the more complete 10 800 spot array experiments ($n = 2$), including a dye exchange. To provide a measure of the double positives, we determined the number of spots above the threshold in both channels (8846). We then calculated the number of spots below threshold (the failure rate) in both samples (1786). These two values are the concordant pairs. Lastly, the number of spots where either the RNA pool was positive and the genomic DNA negative (71; consisting primarily of tissue-specific genes) or vice versa (97; primarily expressed sequence tags and hypothetical proteins) was determined to provide a measure of the difference in the coverage between the two standards or discordant pairs. This is demonstrated in Figure 1a and magnified in Figure 1b. To test the hypothesis that one method was better than the other, we applied the McNemar test with a

Yate's correction. The McNemar test was chosen because it allows the arrays to be treated as a single experiment, with each spot representing an independent trial ($n = 10800$). With one degree of freedom, both hypotheses were valid with an χ^2 value of 3.72 and $P = 0.0538$, which is well above the 95% confidence level, indicating that both reference standards behaved identically.

The failure or double negative rate is an important measure of the overall performance of the standard and the array. To determine the cause of this failure rate, a survey of each of the spots and the reason for its failure to hybridize was undertaken. First, 127 of the missing spots were negative controls, buffer blanks, and empty wells that were not expected to be above threshold, leaving 1659 spots unidentified. To better determine which spots were bad and which were not hybridizing for other reasons,

array data were gathered from other experiments using the same array print and then analyzed by the same means used for the reference standards. With $n = 5$ including dye reversals and the use of indirect labeling, 427 of the 1659 spots were identified as never being above threshold in any experiment. By using reproducibility as the threshold of success, that number increases to 873 spots if the minimum requirement for inclusion is defined as being above threshold in more than half of the experiments. This accounts for only a portion of the spots with no signal, so further analysis was required.

A test was employed to determine the quality of the spots using labeled clone amplification primers as a third independent reference standard. While the primer does not provide a direct measure of the hybridization process, every spot on the array has a copy of

both primers. Therefore, if a spot fluoresced, then DNA was spotted. The two arrays were hybridized with the primer standard in both channels, and the results were compared to genomic DNA and RNA pool reference standards. Across all four arrays and three standards, only 158 spots did not hybridize above the threshold in any experiment. These results beg the questions: "Is genomic DNA complete in the context of a processed message?" and "Is the RNA pool representative of the total transcriptome?" However, the primer standard, which is complete in the context of the array because it is incorporated into every DNA spot, had 1865 spots that did not exceed the threshold value in any of the four trials. This indicates a problem not in the standardization technique but in the hybridization characteristics of the sequences themselves.

While there is no statistical difference between the two standards, one must consider other factors when selecting a reference standard for a series of array experiments. Possibly the most

relevant factor for consideration is that the RNA pool is likely to continue as the de facto standard for human microarrays. The Stratagene RNA reference pool is already in use by hundreds of laboratories (personal communication) including the Stanford Array group (10). Therefore, the reference pool has already gained wide acceptance. However, the utility of any RNA pool as a standard hinges on the availability, cost, and the lot-to-lot reproducibility. In other organisms where an RNA pool is not available or feasible, genomic DNA will be the standard of choice because it is universally available and, by current definitions, complete.

REFERENCES

1. **Bellenson, J.L.** 2000. Expression data and the bioinformatics challenges, p. 139-165. *In* M. Schena (Ed.), *DNA Microarrays, A Practical Approach*. Oxford University Press, Eynsham, UK.
2. **Eisen, M.B. and P.O. Brown.** 1999. DNA arrays for analysis of gene expression. *Methods Enzymol.* 303:179-205.
3. **Ermolaeva, O., M. Rastogi, K.D. Pruitt, G.D. Schuler, M.L. Bittner, Y. Chen, R. Simon, P. Meltzer, et al.** 1998. Data management and analysis for gene expression arrays. *Nat. Genet.* 20:19-23.
4. **Harrington, C.A., C. Rosenow, and J. Retief.** 2000. Monitoring gene expression using DNA microarrays. *Curr. Opin. Microbiol.* 3:285-291.
5. **Hegde, P., R. Qi, K. Abernathy, C. Gay, S. Dharap, R. Gaspard, J.E. Hughes, E. Snesrud, et al.** 2000. A concise guide to cDNA microarray analysis. *BioTechniques* 29:548-556.
6. **Hughes, T.R., M. Mao, A.R. Jones, J. Burchard, M.J. Marton, K.W. Shannon, S.M. Lefkowitz, M. Ziman, et al.** 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* 19:342-347.
7. **Hughes, T.R., M.J. Marton, A.R. Jones, C.J. Roberts, R. Stoughton, C.D. Armour, H.A. Bennett, E. Coffey, et al.** 2000. Functional discovery via a compendium of expression profiles. *Cell* 102:109-126.
8. **Khan, J., R. Simon, M. Bittner, Y. Chen, S.B. Leighton, T. Pohida, P.D. Smith, Y. Jiang, et al.** 1998. Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.* 58:5009-5013.
9. **Lander, E.S.** 1999. Array of hope. *Nat. Genet.* 21:3-4.
10. **Novoradovskaya, N., T. Payette, A. Novoradovsky, J. Braman, N. Chin, A. Pergamenschikov, M. Fero, and Botstein.** 2000. Pooled reference RNA for human microarrays. *Stratagene Strategies* 13:121-122.
11. **Rosner, B.** 1995. *Fundamentals of Biostatistics*. Wadsworth Publishing Company, Belmont, CA.
12. **Ross, D.T., U. Scherf, M.B. Eisen, C.M. Perou, C. Rees, P. Spellman, V. Iyer, S.S. Jeffrey, et al.** 2000. Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* 24:227-235.
13. **Shibata, D.** 1999. Pattern recognition and arrays: the times are a-changing. *Am. J. Pathol.* 154:979-980.
14. **Talaat, A.M., P. Hunter, and S.A. Johnston.** 2000. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nat. Biotechnol.* 18:679-682.
15. **Talaat, A.M., W. Hale IV, R. Lyons, H.R. Garner, and S.A. Johnston.** Genomic standards for gene expression profiling. (In Press.)
16. **Warrington, J.A., S. Dee, and M. Trulson.** 2000. Large-scale genomic analysis using Affymetrix GeneChip probe arrays, p. 119-148. *In* M. Schena (Ed.), *Microarray Biochip Technology*. Eaton Publishing, Westborough, MA.
17. **Zhou, Y., P. Kalocsai, J. Chen, and S. Shams.** 2000. Information processing issues and solution associated with microarray technology, p. 167-200. *In* M. Schena (Ed.), *Microarray Biochip Technology*. Eaton Publishing, Westborough, MA.

We thank Adel M. Talaat for his help in developing the genomic DNA labeling protocol and the laboratories of Rodger Schultz and John Minna for giving generously of their the DNA and RNA stocks. In addition, we thank Robert Powell and Walker Hale of the UTSW Microarray Core for their assistance with troubleshooting and array analysis and Alexander Pertsemididis for stimulating conversations and his help with preparing the manuscript. This work was funded by the NCI SPORE P50 CA70907 grant for technological development and the Program for Genomic Applications (grant no. 5U01HL6688002) from the National Heart Lung and Blood Institute. M.R.W. and T.M. contributed equally to this work. H.G. is on the Scientific Advisory Board of Stratagene, and our group was involved with testing the Stratagene Human Reference RNA.

Received 18 December 2001; accepted 5 March 2002.

Address correspondence to:

Dr. Harold Garner
5323 Harry Hines
Dallas, TX 75390-8591, USA
e-mail: harold.garner@utsouthwestern.edu

**For reprints of this or
any other article, contact
Reprints@BioTechniques.com**