MICROARRAY Technologies

Research Report

Assessing Sources of Variability in Microarray Gene Expression Data

BioTechniques 33:916-923 (October 2002)

Susan E. Spruill, Jun Lu[1], Sarah Hardy[1], and Bruce Weir[1]
DNA Sciences Laboratories, Morrisville, and [1]North Carolina State University, Raleigh, NC, USA

ABSTRACT

Experiments using microarrays abound in genomic research, yet one factor remains in question. Without replication, how much stock can we put into the findings of microarray experiments? In addition, there is a growing desire to integrate microarray data with other molecular databases. To accomplish this in a scientifically acceptable manner, we must be able to measure the validity and quality of microarray data. Otherwise, it would be the weakest link in any integration process. Validating and evaluating the quality of data requires the ability to determine the reproducibility of results. Data obtained from a microarray experiment designed as a feasibility test provided a unique opportunity to partition and quantify several sources of variation that are likely to be present in most microarray experiments. We use this opportunity to discuss the origins of variability observed in microarray experiments and provide some suggestions for how to minimize or avoid them when designing an experiment.

INTRODUCTION

At the 2000 Critical Assessment of Microarray Data Analysis meeting, keynote speaker John Weinstein remarked that microarray research was unique in that the literature was far ahead of the reality (23). Weinstein felt that the analytical methodologies being applied to microarrays were actively under development while the technology itself was still uncertain and far from stable. As the technology becomes more prevalent, the desire to share microarray data and to integrate these data into other molecular databases increases. Noting this desire, Becker (3) critiqued the current status of microarray publications with regard to validity and quality. He pointed out that the fundamental weakness of any shared data effort is the data quality and the ability to assess data quality independently. He noted that results of single microarray experiments with no replication are often published with little statistical analysis or with simple statements about the representative outcome of the experiments. He also pointed out that there would likely be little agreement on the most appropriate statistical analysis, but there should be agreement on traditional, simple measures of quality and reproducibility. Following a similar theme, several speakers at the 10th Annual Bioinformatics and Genome Research meeting emphasized the importance of performing formal statistical analyses and automating these analyses, thereby making microarray analyses more objective (15).

For unreplicated experiments, cluster analysis has been the most common statistical analysis approach used on microarray data. This is partially due to its simplicity and partially due to the nature of scientific questions. Because there is usually no a priori rationale for defining clusters, clustering methods tend to be hierarchical in nature, grouping genes based on the magnitude of expression, magnitude of differences, or correlations in gene expression (1,4, 5,7,20,21,27). Bayesian approaches to clustering and self-organizing maps remove some of the rigid structure of hierarchical clustering but are still largely exploratory in nature (24).

Without replication, we cannot distinguish true differences in gene expression from those differences merely caused by experimental variability. Previous studies have shown that variability between replicates is often large. Even moderate size differences (2- to 10-fold) can require many replications (6). The difficulty in distinguishing true differences in gene expression caused by experimental variability has been recognized, and yet very little has been done to address this issue (25). Simple rules that eliminate genes with fewer than 2- or 3-fold expression changes completely miss biologically important genes that have small fold changes. However, these genes may be highly significant statistically because they could be measured with high precision as a result of replication (26). Experimental replication is fundamental to reliable scientific discovery in genetic research. Understanding the source of noise in the process, controlling it, and possibly eliminating it are essential to drawing reliable inferences (14).

Baldi and Long (2) argue that the

Student's *t* test is similar to a simple-minded fold change approach because the logarithm of a ratio is the same as the difference of two logarithms. The fundamental problem with a *t* test is that it is dependent on sample size. A Bayesian probabilistic approach introduced by Baldi and Long (2) is apparently more robust when there are fewer replicates. However, the need for replication still exists. Important insights into the nature of inherent variability are obtained from the replication of experiments. Lee et al. (14) showed that a single microarray output is subject to substantial variability, even under relatively controlled conditions of an experiment. The authors investigated the variability associated with unevenness of glass slide surfaces and locations of DNA spots on slides using a single dye experiment, thus evaluating the minimum variability that is likely to be inherent in the system. Still, this variation was considerable. Even within the same array, spots containing the same clone do not produce identical fluorescent signals (9,10). Replication does not ensure duplication of results. This fact cannot be quantified if replication is not used (14). As the number of replicates increases, it is possible to be more confident of the results. Wildsmith et al. (25) found that seven replicates gave acceptable reproducibility, raising doubts about the validity of single or duplicate microarrays for the semi-quantitative analysis of gene expression.

The replication of spot measurements either within or between arrays is essential because significance levels of each gene are determined on the basis of distinct estimates of intra-gene variability (26). However, there are other major sources of variation to consider in microarray processing such as probes, target and array preparation, hybridization processes, background and overshining effects, and image processing (17,22). Systemic variations in pen geometry could result in different amounts of cDNAs being deposited on slides (22). PCR amplification is difficult to quantify or may fail to complete, resulting in different amounts of product (22). Hybridization efficiency is influenced by temperature, time, buffering conditions, and the overall amount of probe (22). Other factors affecting the variability of microarray data include nonlinear transmission characteristics, saturation effects, and variation in spot shape increase the variation from image analysis (22). Newton et al. (18) recommend methods for preprocessing the image data using an empirical Bayesian approach to reduce errors in subsequent analyses. Schuchhardt et al. (22) suggest using refined normalization procedures to remove or minimize systemic error from the data before analysis.

Other reported sources of variability have included variability in RNA isolation, tissue-to-tissue variability, and within-tissue heterogeneity (23). Analysis of variance (ANOVA) of spot intensities has revealed that label type, RNA type, and enzyme type have a great impact on signal intensity (25). Loos et al. (16) measured spot-to-spot variability and found it to be 3.8% between spots on the same slide, 5.0% between spots on different slides, and 8.1% between spots from different label reactions.

As microarray technology advances, we are discovering an increasing number of factors that can contribute to variability in microarray data. The need for a priori experimental design considerations is more important than ever. Kerr and Churchill (11) framed the problem of microarray analysis as one of making estimates of relative expression for genes that are not based on ancillary sources of variation. They also (9,10) set the stage for identifying experimental designs that are applicable in microarray research. The Latin square design (dye-swap) is a particularly efficient design for two-dye experiments. Other designs recommended for consideration include incomplete block designs, A-optimality designs, loop designs, and reference designs (9). All of these designs are conducive to varying forms of ANOVA methods. In an ANOVA framework, one uses the data to estimate both relative expression and the magnitude of noise (8,12). ANOVA methods also have the advantage of being a proven statistical tool in biological sciences, dating back to the early days of modern statistics and experimental design. The more advanced ANOVA methods such as mixed-models (26) are even more powerful and allow for flexibility in estimating multiple fixed and random effects in the models.

The purpose of this paper is to utilize multiple levels of replication in a nested factorial design to assess the sources of variability in data obtained from a microarray experiment. We use ANOVA as previously described (8,12, 26) to partition and quantify known sources of variation. We also describe more effective ways to control for noise through changes in techniques or considerations in experimental design.

MATERIALS AND METHODS

The microarray probes were obtained from human liver cDNA purchased from BD Biosciences Clontech (Palo Alto, CA, USA). Microarray slides were prepared by spotting cDNA from various genes onto glass using Gen II Array Spotter (Amersham Biosciences, Piscataway, NJ, USA). Each probe spot represented a 50-base oligomer for one gene, transferred by a robotic pen from a 96-well microplate onto the glass surface. Two replicate microplates were used in the robotic spotting process to ensure that a sufficient amount of DNA was available for spotting each gene across 30 identical glass slides. Generally, each gene was present in only one well in each plate. However, two genes, glucose-6 phosphate dehydrogenase (G6PD) and β-actin, were replicated in eight wells on each plate. Because of the high degree of replication and for the purposes of illustration, only these genes were used to assess the source of variability in this paper.

Our target samples were obtained from 10 cadaverous livers that were suitable for donor transplantation. Total RNA from was extracted from each liver, labeled with Cy3-dCTP, and hybridized onto three slides, thus creating a nested design of replicated genes within slides, within livers.

DNA from a single well was spotted twice from a single pen dip, providing two replicate spots that originated from a single well and pen combination. Two pens operated in tandem to draw two samples from a plate simultaneously and place them onto the same slide. Each gene was spotted 32 times within a single slide. Both genes were spotted on the same slides.

Raw intensity readings were obtained from ArrayVision™ Software

(Imaging Research, St. Catharines, Ontario, Canada) and were adjusted for regional noise by subtraction. Regional noise was defined as intensity readings obtained from dead space in between spots in each of the four quadrants of a slide. The upper left-hand quadrant encompassed spots contained in columns 1 and 2 coming from pen 1. The upper right hand quadrant encompassed spots also contained in columns 3 and 4 coming from pen 1. The lower left-hand and right-hand quadrants were similarly defined by pen 2. All spot intensities within a quadrant were adjusted by subtracting an average background reading for their respective quadrants. Any readings less than the background were set to zero. These data were then evaluated for deviations from normal assumptions using the visual inspection of Q-Q plots and residual plots. No efforts were made to identify or remove outliers because the purpose of this analysis was to measure the magnitude of variation observed from different sources.

ANOVA (PROC GLM; SAS Institute, Cary, NC, USA) (19) was used to partition the sources of variation within the microarray experiment. The model was based on the following:

$$Y_{ijklmn} = L_i + S(L)_{ij} + G_k + P_l + M_m + G*P_{kl} + G*M_{km} + P*M_{lm} + W(P*M)_{lmn} + G*W(P*M)_{klmn} + L*G_{ik} + L*P_{il} + L*M_{im} + L*G*M_{ikm} + G*S(L)_{ijk} + P*S(L)_{ijl} + M*S(L)_{ijm} + G*P*S(L)_{ijkl} + E_{ijklmn}$$

where Y represents background-adjusted intensity reading; L represents liver (I = 1, 2,…10); S represents slide (j = 1,2,3); G represents gene (k = G6PD, β-actin); P represents pen (l = 1,2); M represents microplate (m = 1,2); W represents wells (n = 1, 2,…8); and E represents residual error. All model terms except for E were treated as fixed effects. As a result, the measure of variation for each term could be evaluated relative to residual error.

Table 1. Sources of Variation from the ANOVA Partition

Source	Degrees of Freedom	Relative Expected Mean Squares
Liver	9	590
Slide (Liver)	20	468
Gene	1	6556
Pen	1	71
Microplate	1	822
Gene × Pen	1	5
Gene × Plate	1	82
Pen × Plate	1	11
Well (Pen × Plate)	12	13
Gene × Well (Pen × Plate)	13	10
Liver × Gene	9	169
Liver × Pen	9	21
Liver × Plate	9	11
Gene × Liver × Plate	9	8
Gene × Slide (Liver)	20	101
Pen × Slide (Liver)	20	11
Plate × Slide (Liver)	20	9
Gene × Plate × Slide (Liver)	20	9
Residual Error	1724	1

RESULTS AND DISCUSSION

Ages of liver donors ranged from 15 to 53 years, with a mean age of 40 years. The samples consisted of eight male donors and two female donors. In addition, four ethnic groups were represented: five Caucasians, two African Americans, two Hispanics, and one Asian. Because of the limited number of liver samples, no attempts were made to adjust the ANOVA for ethnicity, age, or gender.

Observed residuals from the ANOVA model were plotted against the predicted values and the normal order statistics for each observation. The visual inspection of these plots did not reveal any gross departures from normal least squares assumptions, indicating that it was unnecessary to normalize the data

using a log transformation before analysis. This eliminated the concern that the data would be artificially truncated because of the presence of log zero (10).

The ANOVA model was partitioned into sources of variation, as described earlier. Arguably, higher-order interaction terms could be added to the model to partition fully all known sources of variation. However, these interactions were expected to be minor in comparison with the sources already described. When a full model containing all higher-order interactions of the main effect terms was fit, it resulted in a coefficient of determination of 0.965. The reduced model that we have described resulted in a coefficient of determination of 0.941, indicating that only an additional 2.4% of the total variability is unaccounted for in the reduced model. The coefficient of determination is the relative decrease in error due to the presence of the observed variables (13). Terms that were removed from the reduced model were reabsorbed into the residual error, increasing the estimate of error by less than 0.04% for every degree of freedom added back into the error. We did not feel that this increase unduly biased the estimation of residual error; therefore, residual error is still considered to be a good estimate of the uncontrolled experimental error. For the purposes of illustration, expected mean squares were normalized relative to the residual error, such that the result is a variance ratio comparable to the F statistic. This allowed for the direct measure of the relative magnitude of each source of variation with respect to experimental error. The resulting partition is provided in Table 1.

The largest source of variation in this study can be attributed to gene differences. Variation due to gene differences was 6556 times greater than the estimated residual error. The two genes analyzed in this study were considered to be housekeeping genes and were therefore expected to be highly expressed in all livers. β-actin is a structural molecule for eukaryotic cells, and G6PD is involved in carbohydrate metabolism. As such, both genes were expected to be expressed in all livers samples but not necessarily at similar levels. On average, spots from the β-actin gene were approximately 92% brighter than spots from G6PD. Differential intensity readings between genes could be attributed to several factors, including differences in gene product availability (mRNA),

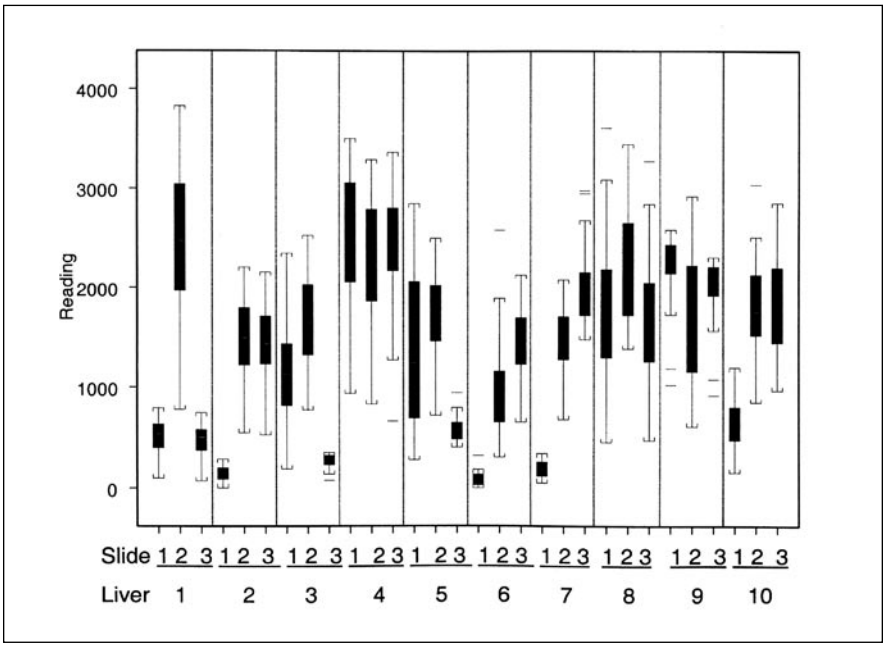

Figure 1. Box plots of the three slides representing each liver. The black boxed area denotes 25th and 75th quartile. The white lines in the black boxes denote the median. The thin lines with caps indicate the upper and lower extreme values within 1.5 times the inter-quartile range. Lines outside of the cap region indicate the outliers.

labeling reactions, or hybridization. In this case, there is no known biological relationship between β-actin and G6PD other than that they are both necessary for cell maintenance.

The liver effects can be defined as the difference in average gene expression, from liver to liver, throughout the experiment. Often in microarray experiments, the treatment difference is derived from tissue from a diseased subject compared to a non-diseased subject. We expect some inherent liver differences to exist, given that the samples came from varying genders, ethnicities, and ages. In our experiment, variation due to liver differences was 590 times greater than the estimated residual error. However, should we expect such large differences to be due to inherent variation alone? More importantly, how would we separate treatment differences from inherent differences if they are confounded with sample origin? In a case-control experiment, this problem could be minimized by sampling diseased and non-diseased tissue from the same individual, thus eliminating factors such as differing age, gender, and ethnicity. However, in studies where the treatment comparison does not allow for sampling the same individual under different conditions (i.e., leukemia vs. healthy), it will be impossible to avoid confounding treatment differences with other factors. In our analysis, we did not attempt to adjust for variation due to gender, age, or ethnic differences in the liver samples because the sample size did not permit this. In addition, since we used a single dye experiment, only one liver sample was hybridized on any given slide. An experiment could be designed to account for sample differences by utilizing two dyes and the more elegant loop design as previously described (9).

Three slides were used to replicate liver effects at the experimental level. In our analysis, slide differences within the liver were nearly 470 times greater than residual error. Recall that all 30 slides in this experiment were identically spotted using the same cDNA samples. We would expect relatively small variation among three slides targeted with the same liver sample primarily because of differences in hybridization time (22). However, in our analysis, the variation among slides within livers was nearly as large as the variation among different liver samples. Figure 1 illustrates the variation among slides within each liver for the G6PD gene. The average intensity readings for slides targeted with the same liver sample varied by as little as 5% and by as much as 1200%, considerably greater than previously observed (16). Further investigation into why slides differed by so much in this study revealed that a separate labeling reaction had been performed for each slide's target material. As a result, the slide differences were also confounded with the labeling reaction differences. In addition, the hybridizations were not performed on the same days. Although this was an unfortunate oversight on the part of the laboratory technician preparing the targets, it illustrates how easily a simple replication can be unduly confounded, which artificially increases variation in the experiment. It would have been preferable to perform only a single labeling reaction for target material to be used on all three slides replicating a liver. However, if it is impossible to produce sufficient target material for three slides with one labeling reaction, the alternate solution would be to pool the material from different labeling reactions before applying the targets to the slides. Likewise, it is important that conditions during hybridization be kept constant. Performing hybridizations over multiple days increases the risk of introducing more variation at this level.

As previously stated, genes were replicated eight times within each slide. However, recall that the spots originated from multiple wells within two microplates. In addition, two pens were used to dispense the spots onto each slide. As a result, only two spots on each slide are true replicates originating from the same sources of variation. Figure 2 shows the box plots of each well, microplate, and pen combination for the three slides representing a single liver and illustrates the large variability within slides. Our ANOVA method revealed that variation attributed to microplates was 822 times greater than residual error. This was the second largest source of variation after gene differences. The microplates were prepared simultaneously into aliquots from the same cDNA sources. For all intents and purposes, the two microplates should be considered true replicates at the sample level. However, our results showed that there was a significant difference between the two plates. The plate replication is further complicated by the fact that two pens are operating on one microplate at the same time. Both pens draw a sample from different wells within the same plate, such that samples in wells 1–4 are picked up by one pen and samples in wells 5–8 are

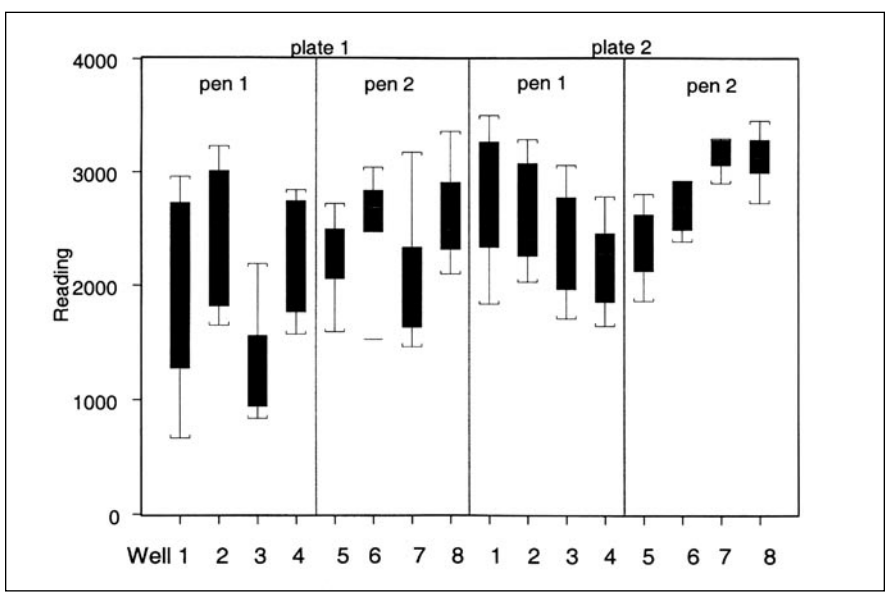

Figure 2. Box plot of the wells within each pen and plate arrangement. This illustration is for G6PD in Liver #3. See Figure 1 for details.

picked up by the other pen. These samples are then deposited simultaneously onto a glass slide. Each pen lays down two spots before returning to the microplate for a new sample, thus creating the only true replicates in this experiment. Considering the relatively small magnitude of the residual error, the variation attributed to the true replication of spots is very small. This finding is consistent with Loos et al. (16), who reported 3.8% differences between replicate spots on the same slide.

Variation attributed to pen difference was 71 times greater than residual error. Compared to other sources of variation, differences attributed to the pens were relatively small but significant. On average, intensity readings differed by only 7% between pens, which was comparable with the average variability between spots from separate hybridizations and separate labeling reactions observed previously (16). However, the microplates were more variable. Average intensity readings from the second microplate were 26% brighter than the readings from the first plate. Evaporation is the only explanation we could find for why microplates would be so significantly different from each other. During the spotting process, both microplates were placed on the platform and made available to the robotic pens. However, all samples were exhausted from the first plate before the pens started to draw samples from the second plate. This allowed extra time for the wells to begin to evaporate, which resulted in higher concentrations of DNA sample in the wells. The potential for the evaporation of sample material in the microplates was unaccounted for when making the glass slides. Clearly, this source of variation will need to be controlled when multiple plates are used. If sufficient sample can be made to fit into a single plate, then plate differences will not be an issue. This solution will likely be unrealistic, given that most robotic systems are designed to handle many plates. For instance, our spotter is designed to hold up to eight plates. In addition, we have no measure for how much drying affects differences in spots printed earlier on the slide as compared to later. When larger numbers of genes are to be spotted, more wells will be needed to hold DNA samples. The ability to program the robotic arms to sample among plates, either at random or in a controlled pattern, will be dependent on the hardware specifications of the spotters. Having the flexibility to sample across plates will be crucial if the plates contain different genes. Because this is a real concern, it would be prudent to avoid confounding microplates with genes. At a minimum, multiple plates should contain incomplete replicates of each gene to allow for gene by microplate variation to be measured and removed from residual error.

Variation among the wells within a plate by pen combination was also relatively small. Our analysis indicated that variation attributed to various wells was only 13 times larger than residual error. In comparison with other sources of variation, this was little more than noise and would not warrant tighter controls when designing an experiment.

Significant interactions usually included genes or plates as a factor. Among the significant interactions were liver and gene. If we had designed an experiment in which livers represented specific treatments (i.e., diseased vs. healthy), then the significance of this interaction would be the most critical component of the analysis because it would indicate that genes were being differentially expressed for different treatments. Testing the effect of the gene by treatment interaction is the primary hypothesis of many microarray experiments, so it is imperative that it be preserved (9,26). The significant variation attributed to a gene by microplate interaction is further evidence that the difference in microplates has an impact on increasing variation at multiple levels. Readings for β-actin were 93% greater than G6PD in spots coming from the first microplate and 91% greater in the second microplate. This slight change in relative difference accounts for the significant interaction and indicates how factors in the laboratory environment such as evaporation could confound readings. Because of the great variation attributed to slide differences, it was also not surprising to find that interactions including slide factors were relatively large. Most prominent among the slide interactions were those that included liver and gene samples. Whereas a liver by gene interaction is desirable, the increased variation attributed to slides would make the determination of significant liver by gene interactions more difficult.

CONCLUSIONS

With the advent of any new technology, developers assume, or perhaps hope, that there is little or no noise in the process and any that does exist is likely to be so minor that it is inconsequential. This assumption would not be a problem as long as the differences needed to answer our scientific questions are large enough to be inherently obvious. However, when we begin to pose scientific questions so refined that the answers require quantifying subtle changes in outcome, we must be willing to invest more effort into teasing out and controlling the sources of variation we were willing to assume did not exist. For example, this study showed that a large amount of variation can be attributed to slide and plate. This variation can mask the true treatment differences in gene expression, increasing both false negatives and false positives if one does not account for these effects. The acceptance and use of microarray technology have matured sufficiently to warrant imposing some new ground rules for acceptable microarray data. It is time to begin expecting sound experimental design strategies when planning microarray studies. Most statisticians would agree that proven statistical methods that are applicable to microarray data are abundant and relatively adaptable to this technology. The problems facing the analysis of microarray data are age old, as is the solution—replication. However, understanding potential sources of variation is not solely the burden of the scientists conducting the experiments. Statisticians should be willing to invest time in the laboratories to evaluate potential sources of variation and to design experiments that evaluate and quantify the impact of these sources.

Our experiment was originally designed as a feasibility study to evaluate the complexity of spotting slides and preparing target materials. The initial evaluation of resulting spot intensities

revealed that too many factors appeared to impact the reliability of the results. Visual inspection of the slides was sufficient to determine that replicate slides were in fact quite different from each other. However, we were fortunate that these data were made available for analysis. Too often, such "dry runs" are given a cursory evaluation and then archived without further analysis or documentation of what did or did not work. As a result, other laboratories are destined to repeat the same type of experiments only to discover similar failings. Still others may charge ahead with experiments, oblivious to the potential sources of variation creeping into their data.

Our analysis of these data provides an example of what sources of variation exist in a typical microarray experiment. We are certain that we have not covered all possible sources. Most obvious is our failure to include any testing for variation attributed to image processing differences. Unfortunately, we did not feel that adequate data regarding variation in the image processing parameters had been collected. There is no doubt that overshining, varying spot shapes, and varying saturation parameters will cause variation in readings. However, we feel that the sources described in our model are among the most prevalent and easily controlled in a microarray laboratory setting. The truth is out there. Replication is a key component to finding it.

ACKNOWLEDGMENTS

Microarray data were obtained from a project funded by DNA Sciences Laboratories (formerly PPGx, a wholly owned subsidiary of DNA Sciences). The analysis was part of a collaboration with PPGx and North Carolina State University's Bioinformatics Program. The authors would like to acknowledge the support of Marco Guida, whose laboratory conducted the microarray experiments yielding these data.

REFERENCES

1. **Alon, U., N. Barkai, D.A. Notterman, K. Gish, S. Ybarra, D. Mack, and A.J. Levine.** 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA *96*:6745-6750.
2. **Baldi, P. and A.D. Long.** 2001. A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. Bioinformatics *17*:509-519.
3. **Becker, K.G.** 2001. The sharing of cDNA microarray data. Nat. Rev. Neurosci. *2*:438-440.
4. **Brazma, A. and J. Vilo.** 2000. Gene expression data analysis. FEBS Lett. *480*:17-24.
5. **Getz, G., E. Levine, and E. Domany.** 2000. Coupled two-way clustering analysis of gene microarray data. Proc. Natl. Acad. Sci. USA *97*:12079-12084.
6. **Hilsenbeck, S.G., W.E. Friedrichs, R. Schiff, P. O'Connell, R.K. Hansen, C.K. Osborne, and S.A.W. Fuqua.** 1999. Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. J. Natl. Cancer Inst. *95*:453-459.
7. **Kaminski, N., J.D. Allard, J.F. Pittet, F. Zuo, M.J.D. Griffith, D. Morris, X. Huang, D. Sheppard, et al.** 2000. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. Proc. Natl. Acad. Sci. USA *97*:1778-1783.
8. **Kerr, M.K., C.A. Afshari, L. Bennett, P. Bushel, J. Martinez, N.J. Walker, and G.A. Churchill.** 2002. Statistical analysis of a gene expression microarray experiment with replication. Statistica Sinica *12*:203-217.
9. **Kerr, M.K. and G.A. Churchill.** 2001. Experimental design for gene expression microarrays. Biostatistics *2*:183-201.
10. **Kerr, M.K. and G.A. Churchill.** 2001. Statistical design and the analysis of gene expression microarray data. Genet. Res. *77*:123-128.
11. **Kerr, M.K., M. Martin, and G.A. Churchill.** 2001. Analysis of variance for gene expression microarray data. J. Comput. Bio. *7*:819-837.
12. **Kerr, M.K., E.H. Leiter, L. Picard, and G.A. Churchill.** 2001. Analysis of a designed microarray experiment. Proceedings of the IEEE-Eurasip Nonlinear Signal and Image Processing Workshop. June 3–6, 2001.
13. **Kim, S., E.R. Dougherty, M.L. Bittner, Y. Chen, K. Sivakumar, P. Meltzer, and J.M. Trent.** 2000. General nonlinear framework for the analysis of gene interaction via multivariate expression arrays. J. Biomed. Opt. *5*:411-424.
14. **Lee, M.T., F.C. Kuo, G.A. Whitmore, and J. Sklar.** 2000. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc. Natl. Acad. Sci. USA *97*:9834-9839.
15. **Leung, Y.F., D.S.C. Lam, and C.P. Pang.** 2001. The miracle of microarray data analysis. Genome Biol. *2*:4021.1-4021.2.
16. **Loos, A., C. Glanemann, L.B. Willis, X.M. O'Brien, P.A. Lessard, R. Gerstmeir, S. Guillouet, and A.J. Sinskey.** 2001. Development and validation of corynebacterium DNA microarrays. Appl. Environ. Microbiol. *67*:2310-2318.
17. **Mirnics, K.** 2001. Mircoarrays in brain research: the good, the bad, and the ugly. Nat. Rev. Neursci. *2*:444-447.
18. **Newton, M.A., C.M. Kendziorski, C.S. Richmond, F.R. Blattner, and K.W. Tsui.** 2001. On differential variability of expression ratios: improving statistical inference about gene expression changes from mircoarray data. J. Comput. Biol. *8*:37-52.
19. **SAS Institute.** SAS/STAT User's Guide, 4th edition. 1990. Version 6, volume 2. SAS Institute, Cary, NC.
20. **Scherf, U., D.T. Ross, M. Waltham, L.H. Smith, J.K. Lee, L. Tanabe, K.W. Kohn, W.C. Reinhold, et al.** 2000. A gene expression database for the molecular pharmacology of cancer. Nat. Genet. *24*:236-244.
21. **Sherlock, G.** 2000. Analysis of large-scale expression data. Curr. Opin. Immunol. *12*:201-205.
22. **Schuchhardt, J., D. Beule, A. Malik, E. Wolski, H. Eickhoff, H. Lehrach, and H. Herzel.** 2000. Normalization strategies for cDNA microarrays. Nucleic Acids Res. *28*:e47i-v.
23. **Siedow, J.N.** 2001. Making sense of microarrays. Genome Biol. *2*:4003.1-4003.2.
24. **Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E.S. Lander, and T.R. Golub.** 1999. Interpreting patterns of gene expression with self-organizing maps: methods and applications to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA *96*:2907-2912.
25. **Wildsmith, S.E., G.E. Archer, A.J. Winkley, P.W. Lane, and P.J. Bugelski.** 2001. Maximization of signal derived from cDNA microarrays. BioTechinques *30*:202-208.
26. **Wolfinger, R.D., G. Gibson, E.D. Wolfinger, L. Bennett, H. Hamadeh, P. Bushel, C. Afshari, and R.S. Paules.** 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J. Comput. Biol. *8*:625-637.
27. **Zhu, J. and M.Q. Zhang.** 1999. Cluster, function, and promoter: analysis of yeast expression array. Pac. Symp. Biocomput. *5*:476-487.

Received 29 March 2002; accepted 5 June 2002.

Address correspondence to:

Susan E. Spruill
Senior Director of Biostatistics
POZEN
1414 Raleigh Road
Chapel Hill, NC 27517, USA
e-mail: sspruill@pozen.com