

Ten Pitfalls of Microarray Analysis

Scott J. Vacha, Ph.D.

Applications Scientist Agilent Technologies

INTRODUCTION

DNA microarray technology is a powerful research tool that enables the global measurement of transcriptional changes between paired RNA samples. While strong biological inferences can be made from microarray transcriptional data, they must be made within the proper biological and experimental context. This is because DNA microarrays capture a static view of a dynamic molecular event. This static view challenges researchers to tease out meaningful biological changes from the associated noise (due to sample acquisition, target labeling, microarray processing, etc.).

There are ten common pitfalls that new users often face when interpreting DNA microarray results. The early recognition of these pitfalls will help minimize false leads and maximize the biological value of the resulting microarray data.





PITFALL 1: FOCUSING ON IMAGE QUALITY OVER DATA QUALITY

The visual inspection of microarray images is an important component to assessing hybridization quality. Often, non-random hybridization issues that may compromise data quality can be flagged upfront by visual inspection (i.e. non-specific binding, scratches, bubbles, etc.).

However, image quality should not be the sole metric for hybridization performance when optimizing DNA microarray experiments. This pitfall is highlighted in **Figure 1.** Here, a scientist performed a cDNA microarray hybridization (**Figure 1a**) and decided to make protocol modifications in order to increase the overall microarray intensity. Although the new modifications did improve the overall intensity (**Figure 1b**) they also introduced additional data skewing that would surely compromise downstream interpretation.

There may also be occasions (although less frequent) whereby poor quality images could hide good data quality. For example, incorrect computer image contrast settings may give the impression that a microarray is of poor quality when, in fact, the underlying data may be sound. So while image quality often correlates with data quality, it should not be used independently as the sole metric of hybridization success.



Figure 1: a) Customer-generated cDNA hybridization results following Agilent's recommended protocols for target labeling. b) The image and scatter plot resulting from several protocol changes that were made to increase microarray intensity. The additional noise in the scatter plot is a direct result of the protocol modifications.



This topic is a continuation of the first pitfall and refers to both microarray and scanner issues. When working with DNA microarrays, users should recognize that signal-to-noise is *more* important than absolute signal.

In terms of microarray signal, there are many factors that can impact spot signal intensity including: target quality, feature diameter, probe quantity, and target specific activity. However, increased spot signal is not an advantage if the corresponding background signal is also increased (**Figure 2**). This is because increased background signal will directly impact detection sensitivity and the ability to extract data from biologically relevant transcripts of low abundance.

This is also an issue regarding scanner PMT adjustments. While researchers may increase PMT gain in order to increase microarray signal, the adjustment could have the detrimental effect of increasing the background signal in a proportionate fashion. **Figure 2** highlights the impact of scanner signal-to-noise on detection sensitivity. For these reasons, users should focus on signal-to-noise rather than absolute signal.



Figure 2: Signal-to-noise. a) This plot represents an intensity histogram of a row of microarray spots. Each peak represents the intensity value of a given spot (red and green channel) and higher peaks correspond to increased spot signal intensity. If the background noise were increased detection sensitivity would be compromised, as several spots would be indistinguishable from the noise. b) This figure plots the green signal-to-noise versus signal for a microarray scanned on three scanners. In this example, only 158 spots are below the detection limit of the Agilent DNA microarray scanner (S/N=3) compared to 8154 spots on Brand Z¹.



PITFALL 3: FAILING TO INTERPRET REPLICATE RESULTS WITHIN THE CONTEXT OF THE 'LEVEL' OF REPLICATION

Although replicate experiments offer statistical benefits, users should be aware that different levels of replication provide different answers. **Figure 3** highlights the main levels of replication in a microarray experiment. In this example one could focus on three mice, three tissue isolates from the same mouse, three RNA isolations, three labeling reactions, three microarrays, or three replicate spots on a single microarray. While all of these are examples of replication, the *level* of replication will directly affect the variability of the replicate measurements.

For example, the red lines in **Figure 3a** represent the use of a single target (from one RNA isolation, of one tissue isolate, of one mouse) for three identical microarray hybridizations. Because the target for all three hybridizations is identical, the only newly introduced variability is a function of the microarray hybridization. This variability includes microarray manufacturing, hybridization, slide washing, and handling. Microarray replicates of the same target will not offer additional information regarding the variability across mice. Similarly, replicates at the level of tissue isolation may include variability due to dissection methods, cellular heterogeneity, and spatial/temporal expression patterns, but will not provide additional information regarding variability across mice.

Finally, replication at the level of biology **(Figure 3b)** will include variability due to strain, disease states, treatment variation, and environmental factors in addition to the variability introduced by subsequent levels (RNA isolation, labeling, etc.). So while technical replicates are important to ensure the procedure is running properly, biological replicates are important to enable transcriptional measurements to be generalized across biological samples (mice in this example). In general, we recommend an appropriate mix of technical and biological replicates depending upon the experimental aim and sample/budgetary limitations.



Figure 3: Levels of replication. a) Replication at the level of microarray hybridization, representing one target hybridized on three identical microarrays. b) Replication at the level of biology, representing the use of labeled target from three independent mice. Details explained in text.



PITFALL 4: ASSUMING THAT STATISTICAL SIGNIFICANCE IS EQUIVALENT TO BIOLOGICAL SIGNIFICANCE

DNA microarray results generally include transcripts that are considered 'differentially expressed' between RNA samples and transcripts that are considered 'unchanged' between samples. This distinction is typically determined by a statistical test using the spot intensity values for each transcript. Although the use of statistics is recommended for defining differential expression, the statistical confidence of differential expression calls may not always reflect the biological relevance of a particular transcriptional change.

This is because all transcripts have a normal physiological range of abundance *in vivo*. Some reports suggest that transcript levels may fluctuate 20-30% in normal biology, simply as a function of how tightly genes are regulated². So although the expression results of a given microarray hybridization may indicate the down-regulation of a particular gene, this change may be well within the normal physiological range for this transcript (**Figure 4a**). As a result, this apparent 'down-regulation' may have less biological significance within the context of the experiment. In order to better estimate the physiological range of transcripts, some researchers perform baseline experiments to model normal transcriptional variation³.

In addition, technical replicates at lower levels (**Figure 4b**) may increase the statistical confidence in a particular measurement by decreasing the variability introduced by replicate hybridizations. In other words, replicate hybridizations from a single RNA prep may increase the statistical confidence of transcriptional changes observed for that RNA sample only. However, this one RNA isolation may not be representative of the gene's biological response across mice, for example. So statistical significance does not always equal biological significance.



Figure 4: a) Each red spot represents the transcriptional abundance of a specific gene with an associated error bar. This figure suggests that the 'experimental' measurement for this gene is within its normal range of transcriptional physiology. b) Replicates at the level of microarray hybridization (see text).



PITFALL 5: IGNORING EXPERIMENTAL DESIGN CONSIDERATIONS

While many researchers appropriately focus on the handling of DNA microarrays, users should also recognize that the *design* of the experiment is as important as the *implementation* of the experiment. For two-color microarray experiments this includes both hybridization design and processing design.

Hybridization design refers to the combination of samples that are compared on a single microarray slide, with one sample labeled green and the other labeled red. Consideration for hybridization design is important because it minimizes the inefficient use of resources (samples, microarrays) and the potential generation of data that cannot appropriately answer the biological question being asked.

Consider the two hybridization designs shown in **Figure 5**, representing the comparison of normal versus treated mice. While the hybridization design of Figure 5a offers the direct comparison of normal and treated mice on the same microarray, the poolednormal sample precludes any measurement of biological variability across normal mice. As a result, it will be difficult to separate biological noise from technical noise in the resulting data. The use of an independent reference sample in **Figure 5b**, on the other hand, will enable the measurement of variability across normal samples but will involve the hybridization to a sample of no biological interest (i.e., the independent reference). While an in depth discussion regarding hybridization design options (direct, loop, balanced block, reference, multi-factorial) is beyond the scope of this paper, researchers should ensure that the hybridization design is consistent with the scientific aim of the experiment.

Processing design refers to the handling of samples throughout the complete course of a microarray experiment (including RNA isolation, target labeling, hybridization, scanning, etc.). The importance of processing design is shown in **Figure 6a**.

Here a scientist hybridized two groups of samples (treated vs. reference; untreated vs. reference) with the expectation that an unsupervised clustering algorithm would discriminate the two treatment groups based upon the similarity of the resulting expression profiles. Unfortunately, because all treated samples were processed on one day and all untreated samples on a different day, the greatest source of variability in the resulting data was the date of processing, rather than the biological treatment group. A more appropriate process design would have included randomizing the samples so that an equal number of treated and untreated samples were processed on the same day (**Figure 6b**).

Although one may be tempted to run a microarray experiment and to figure out the data later, careful up-front consideration must be made to proper hybridization design and processing design. Remember, the design of an experiment is as important as the implementation of the experiment.



Figure 5: Sample hybridization design options for comparing RNA from normal and treated mice. a) This figure represents the separate comparisons of pooled RNA from normal mice versus the RNA from treated mice on three separate microarray hybridizations. b) Each RNA sample (normal or treated) is compared to an independent 'reference' RNA sample, indicated by the blue box.



PITFALL 6: USING EXPERIMENTAL CONDITIONS THAT ARE DIFFERENT FROM THE ERROR MODEL CONDITIONS

The goal of any measurement tool is to provide an estimate of quantitative truth. In the case of DNA microarrays, this 'truth' is of differential gene expression. And, like other quantitative tools, DNA microarray measurements are typically associated with an estimate of measurement error.

For DNA microarrays, this error can be systematic error (i.e., error that can be corrected for such as background subtraction or dye normalization) or random error (i.e., error that cannot be captured, but which can be modeled). To estimate the random error associated with expression measurements from a particular microarray platform, one can perform many hybridizations under identical experimental conditions. With this approach, the normal level of noise associated with a microarray platform can be estimated, independent of biology or technician. Once the random noise associated with a microarray platform is understood, this information can be applied to the interpretation of future, smaller data sets.

While a complete explanation regarding the use of error models is beyond the scope of this paper (for review, see reference 4), the concept behind this pitfall is straightforward. Error models are only accurate under the identical conditions that were used to model the error. In other words, researchers who are implementing an error model in their analyses must use the same experimental conditions that were used for the initial development of the error model. This means using the same labeling protocols, wash conditions, scanner models, and manufacturer's recommendations. Any changes introduced in the process or protocols may add noise that was not initially present in the experiments used to estimate the random error. As a result, the error model (random error estimation) may not be accurate.



Figure 6: Sample results from a poorly processed microarray experiment. a) The unsupervised hierarchical cluster of 9 microarray hybridizations (Normal vs. Reference; Treated vs. Reference) shows grouping of similar experiments based upon the day of processing rather than sample biology. Here processing noise is greater than biological noise. b) An alternative sample processing strategy randomizes the sample handling across two days in order to minimize the processing bias. Here an equal number of untreated samples are processed on Day 1 as treated samples.



PITFALL 7: PAYING MORE ATTENTION TO THE MAGNITUDE OF THE LOG RATIO THAN THE SIGNIFICANCE OF THE LOG RATIO

In early microarray experiments, many researchers filtered microarray data by defining an arbitrary fold-change cutoff for transcripts, such as a 2-fold cutoff. This refers to the ratio of cyanine 5-red labeled target to cyanine 3-green labeled target. If a microarray spot had an intensity of 10,000 counts in the red channel and 5,000 counts in the green channel it would have a ratio of 10,000/5,000= 2, with Log₁₀2=0.3. Because of the poor quality of early microarray technology, only transcripts with a fold change greater than 2 (i.e., x<-2 or x>+2) were considered biologically real. Less attention was given to transcripts with smaller fold changes. Therefore, users focused on the *magnitude* of the Log Ratio in defining 'true' transcriptional differences. This fold-change threshold is depicted in **Figure 7**.

Another approach for defining 'true' transcriptional changes focuses on the *significance* of the Log Ratio. This refers to a statistical definition of significance whereby an estimate is assigned for the probability(P) that a given Log Ratio could occur by chance alone. In other words, if we assume that no transcriptional differences exist between two RNA samples (i.e., average Log Ratio=0), statistics can be used to estimate if an observed change is consistent with this assumption. For example, a Log Ratio with the estimate P<0.01 would suggest that this Log Ratio measurement would be observed about 1% of the time in repeated samplings by chance alone, assuming that its true Log Ratio=0. The smaller the P-value, the less likely the measurement would be observed by chance alone and the more likely that the change is reflective of true differential expression (i.e., the assumption of Log Ratio=0 is not supported). Another way to think of P-value thresholds is the acceptable false-positive rate for a given microarray experiment. So setting a Log Ratio P-value threshold, such as P <0.001, is another approach to filtering microarray data. Only transcripts that pass this filter (i.e., have a P-value equal to or less than 0.001) are considered statistically significant.

As suggested by this pitfall, users should place equal or greater emphasis on the statistical significance associated with Log Ratio values rather than simply the magnitude of the Log Ratio values. The importance of this point is highlighted on the inset of **Figure 7.** Here two transcripts of similar mean intensities are shown with different Log Ratio magnitudes. Because the transcript with the larger fold-change also has a large measurement error, it is not considered statistically significant. In this example, the transcript with the smaller measurement error and Log Ratio magnitude is considered significant.

The arrows in **Figure 7** indicate the difference between the two filtering methods. By applying a 2-fold threshold to these data, the transcripts marked by the thick arrow would be considered false-positives, while the data marked by the thin arrow would be considered false-negatives. So it is as important to consider the statistical significance of Log Ratio measurements when filtering data, as it is to consider the magnitude of the Log Ratio measurements.





Figure 7: Log Ratio vs. Log Intensity plot of two microarray hybridizations, where each dot represents a transcript's error-weighted averaged Log Ratio across two hybridizations. Blue dots represent genes that are not considered statistically significant at P<0.01, red dots represent genes that are significantly up-regulated and green dots are genes that are significantly down-regulated. The dashed line represents a 2-fold Log Ratio threshold.



PITFALL 8: AUTOMATICALLY ASSUMING THAT Q-PCR, RPA, OR NORTHERN BLOT ANALYSIS IS NEEDED TO CONFIRM EVERY MICROARRAY RESULT

In the early application of DNA microarray technology, it was common to confirm observed expression changes by an alternative technology such as Q-PCR, Northern Blots, or Ribonuclease Protection Assays. This confirmation approach was necessary to screen out false-positive results due to the poor quality of early printing methods and the inherent challenges of cDNA clone handling (contamination, PCR issues, re-sequencing, clone tracking/storage, etc.). Today, because of improved manufacturing quality and content quality (*in situ* oligo synthesis, ink jet technology, no cDNA clone handling), the downstream approaches to data confirmation are not strictly limited to these methods. Rather, the confirmation approach should be consistent with the scientific aim of the experiment.

Figure 8 highlights four experimental applications and the alternative confirmation methodologies that may apply. For example, if DNA microarrays are used to suggest a cellular phenotype that discriminates cluster groups of tumors (**Figure 8a**), then the confirmation approach may focus on the hypothesized phenotype rather than confirming the specific transcripts comprising the cluster. For example, Bittner *et. al.* predicted differences in cutaneous melanoma spreading and migration based upon DNA microarray results and confirmed this prediction by a series of cellular assays to measure motility and invasiveness⁵. Similarly, if DNA microarrays are being used to develop a prognostic classifier for metastasis or BRCA1 mutations, then the confirmation approach of the classifier may include testing independent primary tumors or sequencing BRCA1 for putative mutations⁶.

If the scientific aim were to predict a deregulated cellular pathway following experimental treatment, then the downstream confirmation approach might include cellular assays that test the integrity of the suggested pathway, rather than performing Q-PCR of every transcriptional alteration comprising the pathway (Figure 8b). Other experimental aims (Figure 8 c,d) may include functional confirmation and the use of RNAi for target validation.

In summary, the downstream confirmation methods should be consistent with the scientific aim of the experiment. This is not to imply that Q-PCR or similar technologies are no longer of value, but simply to suggest that technological improvements no longer necessitate the confirmation of every transcriptional change in a microarray experiment.



Figure 8: Different experimental goals may necessitate different confirmation methods. The four experimental aims represented here are described in the text.



PITFALL 9: CUTTING UPFRONT COSTS AT THE EXPENSE OF DOWNSTREAM RESULTS

Although focus is often placed on the cost of a microarray slide, this can be insignificant relative to the costs associated with sample acquisition and downstream experimentation (Figure 9). Sample acquisition costs may include obtaining precious tumor biopsies, developing animal knockout models,

synthesizing new compounds, or cloning transfection constructs, for example. These are the costs incurred prior to the microarray hybridization. Downstream costs may include the time/labor/energy involved with interpreting microarray data as well as the resulting experimental steps that are pursued as a direct result of this interpretation.

Since it only takes about three days to perform a microarray hybridization (compared to the weeks or months involved with sample acquisition and data interpretation) it is critical that the microarray results are an accurate reflection of biology and not of poor quality microarrays, inappropriate experimental design, or improper sample handling. The time and costs associated with a microarray experiment are generally lower than the time and costs associated with pursuing poor quality data.

One example of this pitfall was previously shown (**Figure 1**) whereby a scientist modified a labeling protocol in order to increase microarray intensity and to avoid the cost of an optimized commercial labeling kit. A second example involved a customer who substituted Cot-1 DNA blocker that was available in his lab for an empty tube of Cot-1 that was recommended by the microarray vendor. Unfortunately, different Cot-1 preparation methods result in different singleton purity levels that can cross-hybridize to DNA features and interfere with the true Log Ratio measurements. So while both modifications were cost effective and thought to be benign, they had a potentially large impact on the resulting data quality.

This is an important pitfall to consider because many researchers sacrifice the use of quality microarrays, reagents and equipment in an effort to minimize cost. However, by doing so they risk spending months interpreting data that may be less accurate than would have been obtained by a greater investment in the microarray experiment. This investment includes the use of quality microarrays, reagents, and scanner, the rigorous adherence to optimized protocols, and the careful consideration of an experimental design that will maximize data interpretation.

Since it only takes a few days to perform a microarray hybridization, users should invest in this process to maximize the value of the resulting data rather than cutting corners to minimize cost and risk generating data that is less reflective of true biological changes.



Figure 9: Relative costs associated with a microarray experiment.



PITFALL 10: PURSUING ONE PATH IN DATA INTERPRETATION

The proper interpretation of DNA microarray results should always be done within the context of biological information, experimental design, and statistical output, as shown in Figure 10. If pursued independently, each individual path in this figure could result in misleading biological interpretation.

First, supporting biological information (such as the experimental hypothesis, clinical information, literature, etc..) is invaluable for interpreting DNA microarray results. However, this knowledge cannot be the sole framework for interpretation in the absence of proper statistics or experimental design considerations. This can lead to biased conclusions or discounted transcriptional observations that conflict with the initial hypothesis. For example, imagine that a scientist predicted cytoarchitectural changes resulting from a specific drug treatment in culture. If the data were interpreted solely within the context of the initial hypothesis, the scientist might simply look for cytoarchitectural genes in the resulting data and mistakenly overlook other meaningful transcriptional changes. So although the biological context is important, the hypothesis should not bias the interpretation in the absence of statistical methods.

The converse of this is also true. Many microarray facilities employ statisticians to cull microarray data and to identify the relevant transcriptional changes. This is important in order to minimize the pursuit of false leads. However, following this path alone may lead to statistical candidate genes that do not make sense within the context of the experiment (i.e., due to the level of replication, hybridization design, normal physiological range, etc.). So this path should not be pursued independently.

The true path to success in data interpretation is at the interface of the three paths shown in **Figure 10.** Data interpretation must be done within the context of biological information, statistical results, and experimental design. As a result, it is recommended that microarray biologists work very closely with statisticians to ensure that statistical interpretation is consistent with the biological/experimental framework of the project.



SUMMARY

In summary, the power of DNA microarray technology is widely recognized for it's utility in basic research, cancer prognosis, toxicogenomics, and drug discovery. However, it's value as a research tool is dependent upon its proper use and appropriate data interpretation. By recognizing the common pitfalls in data analysis, new users will minimize the time and costs associated with pursuing false leads and maximize the biological meaning present in microarray data sets.

REFERENCES

- 1. http://www.chem.agilent.com/scripts/LiteraturePDF.asp?iWHID=32461
- King, H and Sinha, A. Gene expression profile analysis by DNA microarrays: promise and pitfalls. JAMA. 286, pp. 2280-2288 (2001).
- Hughes, TR., Marton, MJ., Jones, AR., Roberts, CJ., Stoughton, R., Armour, CD., Bennett, HA., Coffey, E., Dai, H., He, YD., Kidd, MJ., King, AM., Meyer., MR., Slade, D., Lum, PY., Stepaniants, SB., Shoemaker, DD., Gachotte, D., Chakruburtty, K., Simon, J., Bard, M. and Friend, SH. Functional discovery via a compendium of expression profiles. Cell, 202, pp. 109-126. (2000).
- Delenstarr, G., Cattell, H., Connell, S., Dorsel, A., Kincaid, R., Nguyen, K., Sampas, N., Schidel, S., Shannon, K., Tu, A., and Wolber, P. Estimation of the confidence limits of oligonucleotide microarray-based measurements of differential expression. In: Microarrays: Optical Technologies and Informatics. Michael Bittner, Yidong Chen, Andreas Dorsel, Edward Dougherty, Editors, Proceedings of SPIE Vol. 4266; pp. 120-131 (2001).
- Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Alberts, D., Sondak, V., Hayward, N.and Trent, J. Molecular classification of cutaneous malignant melanoma by gene expression pro filing. Nature 406, pp. 536-540 (2000).
- Van't Veer, LJ., Dai, H., Van de Vijver, MJ., He., YD., Hart, A., Mao, M., Peterse, HL., Van der Kooy, K., Marton, MJ., Witteveen, AT., Schreiber, GJ., Kerkhoven, RM., Roberts, C., Linsley, PS., Bernards, R. and Friend, S. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415. pp.530-536. (2002).

Ordering Information

www.agilent.com/chem/dna u.s. and canada 1 800 227 9770 japan 0120 477 111 europe: marcom_center@agilent.com global: dna_microarray@agilent.com

© Agilent Technologies, Inc. 2003

Research Use Only

Information, descriptions and specifications in this publication are subject to change without notice.

Printed in the U.S.A. October 21, 2003 5989-0210EN

