

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2010 April ; 19(4): 1000–1003. doi:10.1158/1055-9965.EPI-10-0044.

Standards affecting the consistency of gene expression arrays in clinical applications

Steven A. Enkemann [Director]

Molecular Genomics Laboratory, H. Lee Moffitt Cancer Center and Research Institute, SRB2, 12902 Magnolia Drive, Tampa, Florida 33612, steven.enkemann@moffitt.org

Abstract

The use of microarray technology to measure gene expression has created optimism for the feasibility of using molecular assessments of tumors routinely in the clinical management of cancer. Gene expression arrays have been pioneers in the development of standards; both for research use and now for clinical application. Some of the existing standards have been driven by the early perception that microarray technology was inconsistent and perhaps unreliable. More recent experimentation has shown that reproducible data can be achieved and clinical standards are beginning to emerge. For the transcriptional assessment of tumors this means a system that correctly samples a tumor, isolates RNA and processes this for microarray analysis, evaluates the data, and communicates findings in a consistent and timely fashion. The most important standard is to demonstrate that a clinically important assessment can be made with microarray data. The standards emerging from work on various parts of the entire process can guide the development of a workable system. But, the final standard for each component of the process depends on the accuracy required when the assay becomes part of the clinical routine: a routine that now includes the molecular evaluation of tumors.

Introduction

Scientists have been using hybridization based assays for the quantitation of the messenger RNA amounts since the 1970s (1). The measurement of one gene by Northern blotting quickly shifted to the measurement of several genes in a single sample (2) and by the early 1990s some laboratories were looking at the expression of several hundred genes at once (3-5). In the mid 1990s two reports demonstrated the use of tiny arrays of probes to measure the expression of thousands of genes in a single experiment (4,5). These “micro” arrays quickly became revolutionary, as well as controversial, techniques for the investigation of biological systems. Today, microarrays mass produced by companies can assess nearly the entire complement of genes expressed in an organism in a single assay. This whole transcriptome analysis has generated a lot of optimism about the personalization of medical treatment, particularly with respect to cancer treatment based on the assessment of gene expression in tumors (6-8).

The early years of microarray experimentation yielded novel insights into diseases and biological systems, but also created a perception that the technology might not be reliable. Groups performing similar experiments published different gene lists (9-11). Perhaps contributing to this was the high cost of early microarray experimentation and the fact that limited space in journals restricted the first microarray reports to the most significant observations and only superficial details of the experimental process. For this reason, one of the first “Standards” established for microarray experimentation was the MIAME standards outlined by the Microarray Gene Expression Data society (MGED) (12). The intent of the MIAME standards was to provide enough experimental details to allow researchers to reproduce published microarray based experiments. However, differences in the experimental details did not explain fully the poor reproducibility observed in early experiments. In direct,

laboratory controlled, experimental comparisons many groups had difficulty obtaining reproducible results (13-15). In most of these reports the blame for poor replication was cast on the uniqueness of the arrays, the variability of individuals, and the nuances of the processing or statistical methods used for the microarray process. The net result was that microarrays were gaining a reputation for poor reliability and therefore validation was necessary. These results were interpreted by some that microarray technology might only be useful for rough draft discovery of important genes, not as a definitive assay (16).

Fortunately, a second phase of experimentation supports a more optimistic view of the reproducibility of microarrays technology. The main reason for poor reproducibility in early experiments was poor gene identification. A series of papers showed that probes frequently did not detect the suggested genes (17-20). Better bioinformatics support led to the ability to identify genes based on the sequence of the probes and to account for biological considerations like splice variants, gene families, and pseudogenes. Sequence bases comparisons of array data produced more favorable replication (21,22). Another contributing factor is that the processing of microarray samples is now fairly routine, therefore experienced laboratories can generate more consistent results. Several large multi-laboratory projects directly tested the reproducibility of microarray processing and now suggest that the technology is sufficiently reproducible for clinical applications (23-26).

Basic Standards

During the brief history of microarrays few standards have been established for clinical applications. One reason for this is that microarrays can be used in multiple creative ways and standards established for one application could restrict the suitability of the technology for other experimental purposes. The basic standards for the industry were established early by the companies that market arrays and the general assessment of a hybridized array has been reported in several reviews (27-29). These standards involve checking the quality of the RNA isolated from the samples (30); evaluating the images of the scanned arrays for gross defects (31); evaluating the overall intensity of the sample hybridization to the array (29,32); and some form of global assessment of the probes with detection values above background (29). This last measure is a crude measure of the distribution of gene expression values for detected genes and gives a rough estimate of the similarity of any two samples in the set of samples used for an experiment. These assessments are not really standards, in that they don't define good or bad, but rely on the majority of the samples to set the standard for each experiment and suggest removing the outliers. But cancer is a heterogeneous disease. Even tumors that have the same pathological appearance can be distinctly different in gene expression. Assays that rely on outliers will have great difficulty in separating technological outliers from biological heterogeneity. The basic standards also focus on the identification of grossly inferior or damaged arrays rather than on the identification of subtle flaws that might affect the reproducibility of a single clinical application. Furthermore, these standards only address a small portion of the entire process that would be involved during clinical use (33). Lastly, the tests cannot be performed on a single array as might be necessary in a clinical setting. For these reasons many laboratories have begun to assess additional components of the microarray process in order to establish more standards.

Clinically Directed Standards

The Food and Drug Administration (FDA) began the process with a series of position papers detailing the need for additional standards (34,35). Scientists from the FDA, the National Institute of Standards and Technology, and laboratories in both academia and industry developed a MicroArray Quality Control (MAQC) project that uses reference samples to allow the coordinated comparison of the processing and analysis of microarray data (34,36). This is

a necessary beginning for the establishment of laboratory proficiency in microarray processing. The External RNA Controls Consortium outlined the need for adding RNA controls to the microarray process (37). These supplemental molecules can be used to verify that the microarray process occurred appropriately for each individual sample. Several studies looked at the effect of the overall process on the final quality measures or classification success (24, 38,39). However, the net conclusions of these studies are that reproducibly good microarray data can be generated; they do not establish standards that can be applied to the next batch of arrays. These guidelines are yet to come. As array data accumulates it will eventually be possible to set defined standards based on the cumulative performance of many well processed samples.

The quality of clinical microarrays can be adversely affected long before the processing begins. Studies have shown that the time between surgical removal and sample processing (or storage) can affect gene expression (40,41). The method of sample preservation can influence the integrity of the RNA (42). But the most important factor to consider is the correct sampling of the tumor (28,43,44). Tumors are heterogeneous and it is important to minimize the amount of normal tissue or non-tumorigenic hyperplasia passed on to the microarray process; this ensures that the measured sample represents the in vivo tumor. Therefore, one standard that has not been outwardly stated is that a pathological review of the sample headed for microarray analysis is essential. It is also vital to have a good clinical process for handling the samples to avoid sample mix-ups (45). This means from operating room to pathological review of the tissue to RNA isolation and microarray processing. It has already been demonstrated that the overall process can be performed accurately and produce useful clinical data (38,46). What remains is the standardization of these good clinical practices. A parallel consideration is the documentation of the clinical progress of the patient from whom the sample was collected. One consideration that is lost in the concept of personalized medicine is who will be interpreting the microarray data and what information is returned to the primary physician; when would the analysis be performed with respect to the treatment of a patient; and is this a singular event or will further analysis be required as additional clinical information is acquired. The interpretation of microarray data may also require additional information about the patient (46,47). How the microarray analysis is incorporated into the entire clinical process will define the need for further standards.

Perhaps the biggest concern in the pipeline of tumor assessment by microarrays is the classification process. There are many papers reporting the impact of array platform, processing kit, and data preprocessing methods on the final gene expression values (48). and just as many studies demonstrating how gene selection techniques can influence both the gene list and the classification of tumor samples (49,50). Additionally the nature to the classification algorithm can influence the success rate of a classifier (46,51,52). In all, there are more than 2000 papers published on the molecular classification of cancer using microarray data and in most cases the array type, gene list, and classification techniques are unique. Therefore it is almost impossible to establish standards to manage this process. Fortunately, this is not a concern when the ultimate goal is concerned. Does the classifier work? For medical uses another consideration is whether the classification contributes to the healthcare of the patient? The only real standard is whether the classification process accurately produces a clinically important metric. The FDA has already established this standard in their initial recommendations for the use of gene expression profiling (53). This is the normal expectation for any in vitro diagnostic, which is what the FDA considers all clinical microarray tests. The FDA has further set standards for how the classification accuracy must be demonstrated. Basically, the assays should be validated in an independent population consistent with the defined use of the assay. Metaphorically, it doesn't matter how a device is built; only that it works when used in the normal usage environment.

In developing a clinical application for microarrays it is good to keep this standard in mind. Once the classification can be performed with the best quality possible at all steps in the process it is possible to work back through the process to determine how much noise is acceptable and still yield good classification success. This raises another important consideration; while it is useful to know when to reject poor quality data, it is also important to know how to recapture the necessary information when a sample is rejected. This has consequences for the patient if this means collecting another sample of the tumor. During development of the gene expression based assay one should determine the likelihood that clinically useful data can be successfully obtained from any given patient. This is a standard that has been ignored in most microarray studies so far. If a classifier works well with high quality data, but this is only achieved in 10% of the patients, the test is not likely to be well regarded. Conversely, if a classifier can work well with marginal quality data that is achievable in 95% of the patients, it is more likely to be widely implemented. Therefore, during development of a clinical assay one should define how frequently good data is acquired in the average hospital. This will take a lot of samples to establish creating plenty of opportunity to define standards for each step of the entire clinical process.

Other uses for microarrays

Microarrays have given true confidence to the concept of personalizing the treatment of cancer patients. The analysis of gene expression is leading the way, but microarrays can be used for many other purposes in the assessment of a cancer patient. Microarrays can be used for genotyping, which will allow the pharmacogenomic or toxicogenomic evaluation of a patient (54). Microarrays can also be used for sequencing to perform a mutational analysis of key oncogenes and tumor suppressors. Arrays can be used to identify amplifications and deletions through comparative genome analysis and assess microRNAs and epigenetic changes that play a role in cancer (55-57). The comprehensive evaluation of a cancer patient will probably involve many of these sources of information. Therefore additional standards will be required for these other applications should a clinical assay emerge. In developing a new clinical assay the most important standards are: Does the assay produce important clinical information and does the assay accurately produce the necessary information? It is also important to consider whether the assay can be performed routinely in a clinical setting. All other standards required of the clinical process will be dictated by the need to make sure these concerns are met.

References

1. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc Natl Acad Sci U S A* 1977;74:5350-4. [PubMed: 414220]
2. Kafatos FC, Jones CW, Efstratiadis A. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res* 1979;7:1541-52. [PubMed: 503860]
3. Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H. Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm Genome* 1992;3:609-19. [PubMed: 1450511]
4. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467-70. [PubMed: 7569999]
5. Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996;14:1675-80. [PubMed: 9634850]
6. Luo J, Isaacs WB, Trent JM, Duggan DJ. Looking beyond morphology: cancer gene expression profiling using DNA microarrays. *Cancer Invest* 2003;21:937-49. [PubMed: 14735697]
7. Yeatman TJ. The future of clinical cancer management: one tumor, one chip. *Am Surg* 2003;69:41-4. [PubMed: 12575779]

8. Mohr S, Leikauf GD, Keith G, Rihn BH. Microarrays as cancer keys: an array of possibilities. *J Clin Oncol* 2002;20:3165–75. [PubMed: 12118031]
9. Tsibris JC, Segars J, Enkemann S, et al. New and old regulators of uterine leiomyoma growth from screening with DNA arrays. *Fertil Steril* 2003;80:279–81. [PubMed: 12909486]
10. Rhodes DR, Barrette TR, Rubin MA, Ghosh D, Chinnaiyan AM. Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 2002;62:4427–33. [PubMed: 12154050]
11. Dooley TP, Reddy SP, Wilborn TW, Davis RL. Biomarkers of human cutaneous squamous cell carcinoma from tissues and cell lines identified by DNA microarrays and qRT-PCR. *Biochem Biophys Res Commun* 2003;306:1026–36. [PubMed: 12821146]
12. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet* 2001;29:365–71. [PubMed: 11726920]
13. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. *Bioinformatics* 2002;18:405–12. [PubMed: 11934739]
14. Tan PK, Downey TJ, Spitznagel EL Jr, et al. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* 2003;31:5676–84. [PubMed: 14500831]
15. Woo Y, Affourtit J, Daigle S, et al. A comparison of cDNA, oligonucleotide, and Affymetrix GeneChip gene expression microarray platforms. *J Biomol Tech* 2004;15:276–84. [PubMed: 15585824]
16. Slonim DK. From patterns to pathways: gene expression data analysis comes of age. *Nat Genet* 2002;32(Suppl):502–8. [PubMed: 12454645]
17. Harbig J, Sprinkle R, Enkemann SA. A sequence-based identification of the genes detected by probesets on the Affymetrix U133 plus 2.0 array. *Nucleic Acids Res* 2005;33:e31. [PubMed: 15722477]
18. Gautier L, Moller M, Friis-Hansen L, Knudsen S. Alternative mapping of probes to genes for Affymetrix chips. *BMC Bioinformatics* 2004;5:111. [PubMed: 15310390]
19. Knight J. When the chips are down. *Nature* 2001;410:860–1. [PubMed: 11309581]
20. Kothapalli R, Yoder SJ, Mane S, Loughran TP Jr. Microarray results: how accurate are they? *BMC Bioinformatics* 2002;3:22. [PubMed: 12194703]
21. Hwang KB, Kong SW, Greenberg SA, Park PJ. Combining gene expression data from different generations of oligonucleotide arrays. *BMC Bioinformatics* 2004;5:159. [PubMed: 15504239]
22. Mecham BH, Klus GT, Strovel J, et al. Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. *Nucleic Acids Res* 2004;32:e74. [PubMed: 15161944]
23. Bammler T, Beyer RP, Bhattacharya S, et al. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods* 2005;2:351–6. [PubMed: 15846362]
24. Dobbin KK, Beer DG, Meyerson M, et al. Interlaboratory comparability study of cancer gene expression analysis using oligonucleotide microarrays. *Clin Cancer Res* 2005;11:565–72. [PubMed: 15701842]
25. Irizarry RA, Warren D, Spencer F, et al. Multiple-laboratory comparison of microarray platforms. *Nat Methods* 2005;2:345–50. [PubMed: 15846361]
26. Shi L, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;24:1151–61. [PubMed: 16964229]
27. Schinke-Braun M, Couget JA. Expression profiling using affymetrix genechip probe arrays. *Methods Mol Biol* 2007;366:13–40. [PubMed: 17568117]
28. Group TABPW. Expression profiling--best practices for data generation and interpretation in clinical trials. *Nat Rev Genet* 2004;5:229–37. [PubMed: 14970825]
29. Heber S, Sick B. Quality assessment of Affymetrix GeneChip data. *OMICS* 2006;10:358–68. [PubMed: 17069513]
30. Auer H, Lyianarachchi S, Newsom D, Klisovic MI, Marcucci G, Kornacker K. Chipping away at the chip bias: RNA degradation in microarray analysis. *Nat Genet* 2003;35:292–3. [PubMed: 14647279]

31. Arteaga-Salas JM, Zuzan H, Langdon WB, Upton GJ, Harrison AP. An overview of image-processing methods for Affymetrix GeneChips. *Brief Bioinform* 2008;9:25–33. [PubMed: 18057073]
32. Gregory Alvord W, Roayaei JA, Quinones OA, Schneider KT. A microarray analysis for differential gene expression in the soybean genome using Bioconductor and R. *Brief Bioinform* 2007;8:415–31. [PubMed: 17906332]
33. Salit M. Standards in gene expression microarray experiments. *Methods Enzymol* 2006;411:63–78. [PubMed: 16939786]
34. Shi L, Perkins RG, Fang H, Tong W. Reproducible and reliable microarray results through quality control: good laboratory proficiency and appropriate data analysis practices are essential. *Curr Opin Biotechnol* 2008;19:10–8. [PubMed: 18155896]
35. Shi L, Tong W, Goodsaid F, et al. QA/QC: challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Rev Mol Diagn* 2004;4:761–77. [PubMed: 15525219]
36. Shi L, Reid LH, Jones WD, et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* 2006;24:1151–61. [PubMed: 16964229]
37. Baker SC, Bauer SR, Beyer RP, et al. The External RNA Controls Consortium: a progress report. *Nat Methods* 2005;2:731–4. [PubMed: 16179916]
38. Mook S, Bonnefoi H, Pruneri G, et al. Daily clinical practice of fresh tumour tissue freezing and gene expression profiling; logistics pilot study preceding the MINDACT trial. *Eur J Cancer* 2009;45:1201–8. [PubMed: 19232484]
39. Pine PS, Boedigheimer M, Rosenzweig BA, et al. Use of diagnostic accuracy as a metric for evaluating laboratory proficiency with microarray assays using mixed-tissue RNA reference samples. *Pharmacogenomics* 2008;9:1753–63. [PubMed: 19018728]
40. Huang J, Qi R, Quackenbush J, Dauway E, Lazaridis E, Yeatman T. Effects of ischemia on gene expression. *J Surg Res* 2001;99:222–7. [PubMed: 11469890]
41. Spruessel A, Steimann G, Jung M, et al. Tissue ischemia time affects gene and protein expression patterns within minutes following surgical tumor excision. *Biotechniques* 2004;36:1030–7. [PubMed: 15211754]
42. Micke P, Ohshima M, Tahmasebpour S, et al. Biobanking of fresh frozen tissue: RNA is stable in nonfixed surgical specimens. *Lab Invest* 2006;86:202–11. [PubMed: 16402036]
43. Jochumsen KM, Tan Q, Holund B, Kruse TA, Mogensen O. Gene expression in epithelial ovarian cancer: a study of intratumor heterogeneity. *Int J Gynecol Cancer* 2007;17:979–85. [PubMed: 17367315]
44. O'Sullivan M, Budhraja V, Sadovsky Y, Pfeifer JD. Tumor heterogeneity affects the precision of microarray analysis. *Diagn Mol Pathol* 2005;14:65–71. [PubMed: 15905688]
45. Sturgeon CM, Hoffman BR, Chan DW, et al. National Academy of Clinical Biochemistry Laboratory Medicine Practice Guidelines for use of tumor markers in clinical practice: quality requirements. *Clin Chem* 2008;54:e1–e10. [PubMed: 18606634]
46. Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822–7. [PubMed: 18641660]
47. Komori T, Takemasa I, Yamasaki M, et al. Gene expression of colorectal cancer: preoperative genetic diagnosis using endoscopic biopsies. *Int J Oncol* 2008;32:367–75. [PubMed: 18202759]
48. Nygaard V, Hovig E. Options available for profiling small samples: a review of sample amplification technology when combined with microarray profiling. *Nucleic Acids Res* 2006;34:996–1014. [PubMed: 16473852]
49. Vinaya V, Bulsara N, Gadgil CJ, Gadgil M. Comparison of feature selection and classification combinations for cancer classification using microarray data. *Int J Bioinform Res Appl* 2009;5:417–31. [PubMed: 19640829]
50. Abeel T, Helleputte T, Van de Peer Y, Dupont P, Saeys Y. Robust biomarker identification for cancer diagnosis with ensemble feature selection methods. *Bioinformatics*. 2009
51. McLaren CE, Chen WP, Nie K, Su MY. Prediction of malignant breast lesions from MRI features: a comparison of artificial neural network and logistic regression techniques. *Acad Radiol* 2009;16:842–51. [PubMed: 19409817]

52. Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics* 2008;9:319. [PubMed: 18647401]
53. Food and Drug Administration, Guidance for Industry and FDA Staff - Class II Special Controls Guidance Document: Gene Expression Profiling Test System for Breast Cancer Prognosis. Released: May 9, 2007. Available from: <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm079163.htm>
54. Freeman K. Toxicogenomics data: the road to acceptance. *Environ Health Perspect* 2004;112:A678–85. [PubMed: 15345379]
55. Strefford JC, Parker H. Array-based comparative genomic hybridization as a tool for analyzing the leukemia genome. *Methods Mol Biol* 2009;538:151–77. [PubMed: 19277573]
56. Sato F, Tsuchiya S, Terasawa K, Tsujimoto G. Intra-platform repeatability and inter-platform comparability of microRNA microarray technology. *PLoS One* 2009;4:e5540. [PubMed: 19436744]
57. Sarkar D, Parkin R, Wyman S, et al. Quality assessment and data analysis for microRNA expression arrays. *Nucleic Acids Res* 2009;37:e17. [PubMed: 19103660]