Executive Summary

The MicroArray Quality Control (MAQC) Project

Toward Consensus on "Best Practices" for the Generation, Analysis, and Application of Microarray Data in the Discovery, Development, and Review of FDA-regulated Products

Leming Shi
National Center for Toxicological Research, US Food and Drug Administration
3900 NCTR Road, Jefferson, Arkansas 72079, USA
Tel: +1-870-543-7387, Fax: +1-870-543-7854
Leming.Shi@fda.hhs.gov, http://edkb.fda.gov/MAQC/

November 21, 2006

Microarray Technology and the FDA's Critical Path Initiative

On March 16, 2004, the US Food and Drug Administration (FDA) released a report on "*Innovation/Stagnation: Challenge and Opportunity on the Critical Path to New Medical Products*", addressing the recent slowdown in innovative medical products submitted to the FDA for approval. The report described the urgent need to modernize the medical product development process – the Critical Path from bench side to bed side – so that the product development process will be more predictable and efficient (http://www.fda.gov/oc/initiatives/criticalpath/). On March 16, 2006, Secretary Mike Leavitt and Acting Commissioner Andrew von Eschenbach released the Critical Path Opportunities List and Report that provide concrete focus for public and private efforts and investments in new tools that could revolutionize product development. Among the 76 opportunities in fields such as genomics and proteomics, imaging, and bioinformatics, "*Biomarker qualification*" and "*Standards for microarray and proteomics-based identification of biomarkers*" were cited as the top two opportunities.

Microarray technology was identified by the FDA's Critical Path Initiative as a key tool that holds "vast potential" for advancing medical product development and personalized medicine through the identification of biomarkers. However, a gap exists between technologies in use today and the technological levels required for application during product development and regulatory decision making. For example, recent publications have raised concerns about the reliability of microarray technology because of the apparent lack of reproducibility between lists of genes (*i.e.*, potential biomarkers) identified as differentially expressed from similar or identical study designs with different platforms or laboratories[1, 2].

The MicroArray Quality Control (MAQC) Project: A Community-wide Effort in Response to the FDA's Critical Path Initiative

On February 11, 2005, in response to the FDA Critical Path Initiative, scientists at the FDA's National Center for Toxicological Research (NCTR), Jefferson, Arkansas formally launched the MicroArray Quality Control (MAQC) project (http://edkb.fda.gov/MAQC/; FDA/NCTR research protocol number: E0720701) in order to address reliability concerns as well as other performance, standards, quality, and data analysis issues[3]. The first phase of the MAQC project (from February 11, 2005 to September 8, 2006) involved 137 scientists from 51 organizations including the six FDA centers (CBER, CDER, CDRH, CFSAN, CVM, and NCTR), government

agencies (the US Environmental Protection Agency, the National Institutes of Health, and the National Institute of Standards and Technology), manufacturers of microarray platforms and RNA samples, microarray service providers, academic laboratories, and other stakeholders. All MAQC participants freely donated their time and reagents for the completion and analysis of the first phase of the MAQC project.

Major findings of the first phase of the MAQC project were published in six research papers on the September 8, 2006 issue of *Nature Biotechnology*[3-8]. Also published in the same issue included a Foreword by Dr. Daniel Casciano (former FDA/NCTR Director) and Dr. Janet Woodcock (FDA Deputy Commissioner), "*Empowering microarrays in the regulatory setting*"[9], an Editorial[10] from *Nature Biotechnology*, three Commentaries from the FDA[11], the EPA[12], and Stanford University[13], and a Glossary[14]. All the MAQC papers are freely available at *Nature Biotechnology*'s website (http://www.nature.com/nbt/focus/maqc/index.html). In addition, all the MAQC papers were published as a supplement to the Nature Publishing Group in October 2006 and distributed to a wide readership. Data are available through GEO (series accession number: GSE5350), ArrayExpress (accession number: E-TABM-132), ArrayTrack (http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/), and the MAQC website (http://edkb.fda.gov/MAQC/). The MAQC project has attracted international attention as can be seen from the positive reporting by *Cell*[15], *Nature*[16], *Science*[17], *Nature Methods*[18], *Analytical Chemistry*[19], and other scientific publications.

Gene expression data on four titration pools from two distinct, commercially available reference RNA samples (samples A and B) were generated at multiple test sites using a variety of microarray-based and alternative technology platforms. The resulting rich reference data set consists of over 1,300 microarray hybridizations, and additional measurements for over 1,000 genes with alternative technologies such as qPCR. The MAQC project observed high intraplatform reproducibility across test sites, as well as interplatform concordance in terms of genes identified as differentially expressed. Platforms with divergent approaches generated comparable results in terms of differential gene expression. In other words, the differential gene expression patterns reflected the same biology despite differences in platform technology. Similar results were observed from a rat toxicogenomics data set[4], in support of the major findings from data generated from the reference RNA samples.

One important goal of the MAQC study is to assess the best performance achievable with microarray technology under consistent experimental conditions so that end users will be able to judge whether the quality of their microarrays is of comparable performance. In doing so, procedural failures of a laboratory or operator may be identified and corrected before precious study samples are profiled. The commercial availability of the two reference RNA samples coupled with the large reference data sets would also allow for the objective evaluation of new array products, reagents, or protocols.

Several unique features set the MAQC project apart from previous cross-platform comparison studies: (1) the enthusiastic participation of the microarray community in an extraordinary team effort; (2) the scale of the MAQC data set with over 1,300 microarrays from more than 40 test sites and 20 microarray platforms; (3) the large number of additional gene expression measurements with alternative technology platforms; (4) the commercial availability to the community of the same batches of the two reference RNA samples used in the MAQC study for subsequent quality control, performance evaluations, and proficiency testing; (5) the extensive sequence-based mapping of probes across platforms; and (6) last but not least, the

identification of statistical explanations for some misconceptions on the comparability of microarray results.

MAQC Phase I: From Confusion to Consensus on Microarray Data Analysis

A major challenge to the microarray user is the existence of numerous options for analyzing the same data set, which lack adequate scientific vetting of their capabilities, implications, and limitations[16]. There is a pressing need to critically evaluate currently available methods with relevant and objective criteria. For example, reproducibility has seldom been, but in the future should be, used as a critical criterion to judge the performance of data analysis procedures. In addition, several differential gene expression profiling studies have demonstrated that the relative expression measures (*i.e.,* difference in transcript abundance between sample types) are more consistent than the absolute gene expression levels. The MAQC data set is expected to be widely utilized by the community in order to reach and promote consensus on the appropriate analysis of microarray data.

Lists of differentially expressed genes selected solely by a statistical significance measure are irreproducible: The MAQC analyses demonstrated[3, 4] that the apparent lack of reproducibility reported in previous studies using microarray assays[1, 2] was likely caused, at least in part, by the common practice of ranking genes solely by a statistical significance measure, for example, *P* values derived from simple *t*-tests, and selecting differentially expressed genes with a stringent significance threshold. The gene lists in the MAQC study were much more concordant when fold change was used as the ranking criterion. This approach also greatly reduced the impact of different normalization methods. In addition, widely used statistical methods such as SAM did not appear to improve interlaboratory or interplatform reproducibility compared to fold-change ranking. Importantly, non-reproducible gene lists could lead to inconsistent biological interpretations, for example, in terms of enriched GO terms and pathways[4]. Fold-change ranking plus a non-stringent *P*-value cutoff can be used as a baseline practice for generating more reproducible signature gene lists[3, 4].

The effect of data normalization on the stability of lists of differentially expressed gene is greatly reduced when fold change is used for gene selection: Data normalization was identified as a major factor for differences when comparing results and data interpretations performed by VGDS (Voluntary Genomic Data Submission) sponsors and FDA reviewers[11]. It should be noted that, although there are many options for normalizing microarray data, when lists of differentially expressed genes are identified by the ranking of fold change, the results are much less susceptible to the impact of normalization methods. In fact, global scaling methods (*e.g.*, median- or mean-scaling) do not change the relative rank-order of genes based on fold change; they do, however, significantly impact gene ranking by *P*-value[3, 4, 7]. The MAQC results suggest that microarray data analysis for the identification of reproducible lists of differentially expressed genes does not need be as complicated and confusing as it has been practiced, and consensus on data analysis appears to be attainable.

MAQC Phase II: Development and Validation of Predictive Signatures and Classifiers

The MAQC Phase I (MAQC-I) has demonstrated the technical reliability of microarray technology in detecting differential gene expression. However, questions remain regarding the reliability of the technology in clinical applications such as disease diagnostics or prognostics, and for tailored treatment based on gene expression profiles[20, 21, 22, 23]. To investigate the capabilities and limitations of microarray technology in such real-life applications, the MAQC

Phase II (MAQC-II) has been launched to address technical and scientific issues involved in the development and validation of predictive signatures and classifiers. Multiple data sets will be collected and distributed to participating organizations for independent analyses with available algorithms. The resulting classifiers will be evaluated at three different levels: within a single data set via cross-validation, validation across multiple data sets from studies with the same study objectives, and prospective validation with additional data from new samples. It is anticipated that the MAQC project, through the community's active participation, will help develop "best practices" for the generation, analysis, and application of microarray data in the discovery, development, and review of FDA-regulated products.

References

1. Tan, P.K. et al. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res* **31**, 5676-5684 (2003).
2. Marshall, E. Getting the noise out of gene arrays. *Science* **306**, 630-631 (2004).
3. Shi, L. et al. The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat Biotechnol* **24**, 1151-1161 (2006).
4. Guo, L. et al. Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol* **24**, 1162-1169 (2006).
5. Canales, R.D. et al. Evaluation of DNA microarray results with quantitative gene expression platforms. *Nat Biotechnol* **24**, in press (2006).
6. Shippy, R. et al. Using RNA sample titrations to assess microarray platform performance and normalization techniques. *Nat Biotechnol* **24**, 1123-1131 (2006).
7. Patterson, T.A. et al. Performance comparison of one-color and two-color platforms within the Microarray Quality Control (MAQC) project. *Nat Biotechnol* **24**, 1140-1150 (2006).
8. Tong, W. et al. Evaluation of external RNA controls for the assessment of microarray performance. *Nat Biotechnol* **24**, 1132-1139 (2006).
9. Casciano, D.A. & Woodcock, J. Empowering microarrays in the regulatory setting. *Nat Biotechnol* **24**, 1103 (2006).
10. Making the most of microarrays. *Nat Biotechnol* **24**, 1039 (2006).
11. Frueh, F.W. Impact of microarray data quality on genomic data submissions to the FDA. *Nat Biotechnol* **24**, 1105-1107 (2006).
12. Dix, D.J. et al. A framework for the use of genomics data at the EPA. *Nat Biotechnol* **24**, 1108-1111 (2006).
13. Ji, H. & Davis, R.W. Data quality in genomics and microarrays. *Nat Biotechnol* **24**, 1112-1113 (2006).
14. Reid, L.H. & Warrington, J.A. A note on nomenclature. *Nat Biotechnol* **24**, ii (2006).
15. Strauss, E. Arrays of hope. *Cell* **127**, 657-659 (2006).
16. Eisenstein, M. Microarrays: quality control. *Nature* **442**, 1067-1070 (2006).
17. Couzin, J. Genomics. Microarray data reproduced, but some concerns remain. *Science* **313**, 1559 (2006).
18. Kiermer, V. Microarray quality in the spotlight again. *Nat Methods* **3**, 772 (2006).
19. Sage, L. Do microarrays measure up? *Anal Chem* **78**, 7358-7360 (2006).
20. Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* **365**, 488-492 (2005).
21. Ioannidis, J.P. Microarrays and molecular research: noise discovery? *Lancet* **365**, 454-455 (2005).
22. Ein-Dor, L., Zuk, O. & Domany, E. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci U S A* **103**, 5923-5928 (2006).
23. Simon, R. Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. *J Natl Cancer Inst* **98**, 1169-1171 (2006).