This summary should not be cited or quoted for public release until *Nature*'s embargo is lifted at:
1900 London Time (BST) / 1400 US Eastern Time Friday 08 September 2006
0400 Japanese Standard Time / 0600 Australian Eastern Time Saturday 09 September 2006

Summary of the Phase I Results of

The MicroArray Quality Control (MAQC) Project

Toward Consensus on 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 (NCTR)
U.S. Food and Drug Administration (FDA)
3900 NCTR Road, Jefferson, Arkansas 72079, U.S.A.
Tel: +1-870-543-7387, Fax: +1-870-543-7854
Leming.Shi@fda.hhs.gov, http://edkb.fda.gov/MAQC/

September 8, 2006

Microarrays represent a core technology in pharmacogenomics that was identified by the U.S. Food and Drug Administration's (FDA) Critical Path Initiative as a key opportunity for advancing medical product development and personalized medicine. The FDA issued the "*Guidance for Industry: Pharmacogenomic Data Submissions*" to facilitate scientific progress and the use of pharmacogenomic data in drug development and medical diagnostics (http://www.fda.gov/cder/genomics/). However, recent publications have raised concerns about the reliability of the microarray technology because of the apparent lack of reproducibility between lists of genes identified as differentially expressed from similar or identical study designs with different platforms or laboratories.

The MicroArray Quality Control (MAQC) project was initiated by the FDA's National Center for Toxicological Research (NCTR), Jefferson, Arkansas on February 11, 2005 in order to address reliability concerns as well as other performance, quality, and data analysis issues. The first phase of the MAQC project (from Feb-11-2005 to Sep-8-2006) involved 137 scientists from 51 organizations including government agencies (the FDA, the U.S. 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.

Gene expression data on four titration pools from two distinct, commercially available reference RNA samples were generated at multiple test sites using a variety of microarray-based and alternative technology platforms, resulting in a rich dataset with over 1,300 microarray hybridizations. The MAQC project observed intraplatform reproducibility across test sites as well as high interplatform concordance in terms of genes identified as differentially expressed. One major result was that platforms with divergent approaches often generated comparable results of differential gene expression. In other words, the differential gene expression patterns generated were reflective of biology regardless of the differences in technology platforms. Similar results were observed from a rat toxicogenomics dataset, validating the major findings from data generated on reference RNA samples. Findings of the MAQC project were published in a series of articles in *Nature Biotechnology*, September 8, 2006. 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://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/).

Several unique features set the MAQC project apart from previous cross-platform comparison studies: (1) the enthusiastic participation of the community in an extraordinary team effort; (2) the scale of the MAQC dataset 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 of the same batches of the two reference RNA samples used in the MAQC study for subsequent quality control, performance evaluations, and proficiency testing by the community; (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 capabilities of microarray results.

The MAQC analyses demonstrated that the apparent lack of reproducibility reported in previous studies using microarray assays was likely caused 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. Widely used statistical methods did not appear to improve interlaboratory or interplatform reproducibility compared to fold-change ranking. Importantly, non-reproducible gene lists led to inconsistent biological interpretations in terms of enriched Gene Ontology terms and pathways. Fold-change ranking plus a non-stringent *P*-value cutoff can be used as a baseline practice for generating more reproducible signature gene lists. The MAQC results suggest that microarray data analysis for the identification of reproducible, differentially expressed genes need not be as complicated and confusing as it has been practiced.

A major challenge to the microarray user is the existence of numerous options for analyzing the same dataset, which lack adequate scientific vetting of their capabilities, implications, and limitations. 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 dataset is expected to be widely utilized by the community in order to reach and promote consensus on the appropriate analysis of microarray data.

The 5th MAQC project face-to-face meeting, which will be open to the public, will be held at the FDA/NCTR, Jefferson, Arkansas on September 21, 2006 from 8:30 AM CDT to 3:00 PM CDT. The meeting will review the major findings from the first phase and formally launch the second phase of the MAQC project on predictive signatures, classification and modeling for clinical diagnosis and prognosis. More information about the second phase of the MAQC project was announced in a *Federal Register* notice available at http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/docs/FederalRegister_MAQC_FollowUp.pdf. It is our expectation 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.