The MicroArray Quality Control (MAQC) Project

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

Summary of the 5th MAQC Project Meeting, September 21, 2006

National Center for Toxicological Research, FDA, Jefferson, AR

Summary by Leming Shi, September 29, 2006
Leming.Shi@fda.hhs.gov; http://edkb.fda.gov/MAQC/

1. **Overview**: The 5th face-to-face MAQC project meeting was held at the FDA's National Center for Toxicological Research (NCTR), Jefferson, AR on September 21, 2006 (8:30 AM – 3:00 PM); detailed meeting agenda can be found at http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/. A total of 69 on-site participants attended in addition to 11 people who participated by phone. The main objectives of the meeting were: (1) to review Phase I results on microarray technical performance and the expected utility of the results; (2) to kickoff the Phase II effort on predictive signatures, classifiers, and modeling in order to realistically assess the capabilities and limitations of microarray technology in clinical (e.g., diagnostics, prognostics, and individualized therapy) and toxicogenomic applications. Meeting participants expressed strong interests in contributing to the Phase II. The NCTR management team, including Dr. William Slikker, Jr. (Acting Director), reiterated NCTR's commitment to the MAQC project as it moves to Phase II.
2. **FDA Acting Commissioner Addressed the MAQC Meeting**: During his visit to the NCTR on September 21, Dr. von Eschenbach, FDA's Acting Commissioner, was presented with a copy of the September 8th issue of *Nature Biotechnology* that focuses on the MAQC results. Dr. von Eschenbach congratulated the MAQC group for reaching the first milestone with the publication of the Phase I results. He emphasized the extreme importance of appropriately integrating and interpreting complex data from new technologies in medical product development and patient care. "… your effort particularly with regard to microarrays is not only critical and essential to the contribution you make to science and technology, but I want you to know I believe it is critical and essential to the contribution we all want to make to the health and welfare of those patients and the public that is depending upon us, whether it's NCI, or FDA, or any of the organizations, and agencies and institutions that are a part of the effort...".

3. **Review of the MAQC Phase I Results** (chair: Rick Jensen): Following an overview presentation of the MAQC project by Leming Shi (FDA/NCTR), the leading authors of the six MAQC research papers published in *Nature Biotechnology* presented the major findings of the Phase I results. Briefly, the MAQC project observed intra-platform reproducibility across test sites as well as high inter-platform concordance in terms of genes identified as differentially expressed. Microarray platforms with

divergent manufacturing 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. The MAQC analyses demonstrated that the reproducibility reported in previous studies using microarray assays could be significantly improved from that observed by ranking differentially expressed 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. Fold-change ranking plus a non-stringent *P*-value cutoff could be used as a baseline practice for generating more reproducible signature gene lists. The MAQC data were also used to evaluate the comparability between microarrays and quantitative gene expression platforms (Federico Goodsaid), the performance of microarray assays based on external RNA controls (Weida Tong), the impact of normalization methods (Rich Shippy), data consistency between one-color and two-color platforms (Tucker Patterson). Similar results were observed from a rat toxicogenomics data set (Lei Guo/Leming Shi), validating the major findings from data generated using human reference RNA samples (Wendell Jones). The Phase I results demonstrated the achievable performance of microarray technology, supporting wider applications in research that will eventually lead to proper utility in clinical and regulatory contexts.

4. **Lessons Learned**: In a presentation titled "*ROC in 3D: reproducibility as a third dimension beyond specificity and sensitivity in gene selection*", Russ Wolfinger (SAS Institute) emphasized the importance of including reproducibility as an essential and independent criterion in addition to sensitivity and specificity in the identification of differentially expressed genes from microarray studies. Federico Goodsaid (FDA/CDER) described the importance and progress of an FDA effort in developing a "Best Practices" document for Voluntary Genomic Data Submissions (VGDS) to the FDA. The implications of the MAQC Phase I results in the development of the "Best Practices" document were discussed. A workshop co-sponsored by FDA/DIA/PhRMA/BIO on "*Best Practices and Development of Standards for the Submission of Genomic Data to the FDA*" will be held at Washington, DC, Nov. 27-28, 2006 (meeting agenda is attached and updated information can be found at www.diahome.org).

5. **Phase II Kickoff** (chair: Weida Tong): With examples from recently published high-profile papers that questioned the utility of microarrays in clinical applications such as cancer diagnosis and prognosis, Leming Shi (NCTR) illustrated the urgent need of the MAQC Phase II effort on predictive signatures, classification and modeling so that we'll have a much better understanding of the capabilities and limitations of the applications of microarray technology in clinical settings by addressing critical issues including the development and validation of predictive models. From a regulatory perspective, Uwe Scherf (CDRH) discussed what is needed for microarrays to be reliably applied in diagnostics. The kickoff session was then followed by three invited presentations from experts experienced in microarray data analysis. Yudong He (Rosetta Inpharmatics/Merck) presented an overview on data quality control, data analysis, and the challenges in clinical applications regarding reproducibility, specificity and sensitivity. Grier Page (University of Alabama at Birmingham) discussed "*Microarray data analysis: from disarray to consolidation and consensus*",

highlighting the needs for the community to reach consensus on experimental design, data normalization, identification of differentially expressed genes, the development of classifiers for diagnostics and prognostics, and other issues. Rich Simon (NCI/NIH) delivered a thought-provoking speech via telephone on "*Myths about the development and validation of predictive classifiers using gene expression profiles*", pointing out many misunderstandings in the common practices of microarray data analysis including biases in the evaluation of the performance of classifiers and the community's addiction to developing complicated but inadequately validated data analysis methods just for the sake of publications.

6. **Phase II Open Discussions** (co-chair: Wendell Jones, Uwe Scherf, and Russ Wolfinger): Participants extensively discussed issues such as the scope and objectives of Phase II, logistics, intellectual properties, data set nomination, criteria for selecting data sets, and criteria for evaluating signature genes and predictive models. Some participants have already nominated data sets for Phase II to consider; others expressed willingness to contribute tissue samples or to run more arrays when needed. It was agreed that before any decision is made, the MAQC group should conduct a survey of the publicly available data sets or private data sets that could be made available to the MAQC under specific access conditions. Individual organizations are encouraged to provide data sets to the MAQC for analysis in Phase II. It was also agreed that a face-to-face meeting would be desirable to review the data set nominations and to lay out the path ahead.

7. **Three Working Groups**: The following three Working Groups (WGs) were established and will work concurrently during the MAQC Phase II:

   A. ***Clinical WG***, to focus on data sets for clinical applications. Coordinators: *Lajos Pusztai* (M.D. Anderson Cancer Center, lpusztai@mdanderson.org), *Uwe Scherf* (FDA/CDRH, uwe.scherf@fda.hhs.gov), and *Wendell Jones* (Expression Analysis, wjones@expressionanalysis.com).

   B. ***Toxicogenomics WG***, to focus on toxicogenomic applications. Coordinators: *Federico Goodsaid* (FDA/CDER, federico.goodsaid@fda.hhs.gov) and *David Dix* (EPA, dix.david@epa.gov).

   C. ***MAQC Titrations WG***, to focus on the MAQC titration samples (including the MAQC Pilot II data from 13 titration mixtures). Coordinators: *Richard Shippy* (GE Healthcare, richard.shippy@ge.com) *Rick Jensen* (University of Massachusetts Boston, roderick.jensen@umb.edu), and *Russ Wolfinger* (SAS Institute, russ.wolfinger@sas.com).

   Everyone is welcome to join the WGs. If you are interested in contributing to a particular WG, please contact the coordinators of the corresponding WG and cc leming.shi@fda.hhs.gov. Leming Shi will continue to coordinate the overall activities of the MAQC project.

8. **The 6th MAQC Project Meeting, Nov. 29, 2006 (tentative)**: Many MAQC members plan to attend the "*Best Practices*" workshop (see item 4). It has been suggested that we could use Nov. 29 and part of Nov. 28 (3 PM -) to review each WG's progress on data set survey and to draft detailed working plans for the MAQC Phase II. Comments and suggestions for the next face-to-face meeting are welcome.

9. **MAQC at IBC's Discovery-2-Diagnostics Conference**: The MAQC keynote panel presentation (4:00 PM – 5:30 PM, Sept. 25, 2006, Boston, MA) was well received. Thanks to all panel members (Wendell, Damir, Rick, Ernie, and Leming) and MAQC members in the audience for making this a success. It was gratifying to see that organizations (*e.g.*, Agilent and Solexa) have been using the MAQC outcomes (*e.g.*, reference RNA samples and reference data sets) to help develop new products for gene expression profiling.

Participants of the 5th MAQC Project Meeting, September 21, 2006, Jefferson, AR

No.	Name	Organization	No.	Name	Organization
1	Wenjun Bao	SAS Institute	41	Baitang Ning	FDA/NCTR
2	Anne Bergstrom Lucas	Agilent	42	Grier P. Page	University of Alabama
3	Richard Brennan	Iconix	43	Tucker A. Patterson	FDA/NCTR
4	Roger D. Canales	Applied Biosystems	44	Roger Perkins	FDA/NCTR (Z-Tech)
5	James J. Chen	FDA/NCTR	45	Mette A. Peters	Rosetta Biosoftware
6	Tao Chen	FDA/NCTR	46	P. Scott Pine	FDA/CDER
7	Tzu-Ming Chu	SAS Institute	47	Mark Porter	Gene Logic
8	Timothy S. Davison	Asuragen	48	Lajos Pusztai	MD Anderson Cancer Center
9	Thon DeBoer	Agilent	49	Feng Qian	FDA/NCTR (Z-Tech)
10	Robert R. Delongchamp	FDA/NCTR	50	Laura H. Reid	Expression Analysis
11	David J. Dix	EPA	51	Brian Rhees	Roche Molecular Systems
12	Yvonne P. Dragan	FDA/NCTR	52	Uwe Scherf	FDA/CDRH
13	Mike Falduto	GenUs BioSystems	53	Joe Shambaugh	Genedata (USA) Inc.
14	Xiao-hui Fan	FDA/NCTR	54	Leming Shi	FDA/NCTR
15	Hong Fang	FDA/NCTR (Z-Tech)	55	Richard Shippy	GE Healthcare
16	Gavin M. Fischer	Stratagene	56	Dave D. Smith	Luminex
17	Steven D. Flanagan	City of Hope Graduate Sch.	57	Frank Staedtler	Novartis Pharma AG
18	Weigong Ge	FDA/NCTR	58	Hongmei Sun	FDA/NCTR (Z-Tech)
19	Federico M. Goodsaid	FDA/CDER	59	Russell S. Thomas	CIIT Centers for Health Research
20	Lei Guo	FDA/NCTR	60	Karol L. Thompson	FDA/CDER
21	Linda C. Haje	Biotech Consultant	61	Weida Tong	FDA/NCTR
22	Paul K. Haje	TeleChem ArrayIt	62	Christophe Van Huffel	Eppendorf Array Technology
23	Tao Han	FDA/NCTR	63	Theressa Veeneman	Jaden BioScience
24	Stephen C. Harris	FDA/NCTR	64	Stephen J. Walker	Wake Forest University
25	Yudong He	Rosetta Inpharmatics	65	Janet A. Warrington	Affymetrix
26	Huixiao Hong	FDA/NCTR (Z-Tech)	66	James C. Willey	Ohio Medical University
27	Roderick V. Jensen	Univ. of Massachusetts Boston	67	Tanya Willie	Eppendorf Array Technology
28	Charles D. Johnson	Asuragen	68	Russ Wolfinger	SAS Institute
29	Wendell D. Jones	Expression Analysis	69	Qian Xie	FDA/NCTR (Z-Tech)
30	Connie Kohne	Jaden BioScience	***Participants via phone***		
31	David Kohne	Jaden BioScience	70	Shashi Amur	FDA/CDER
32	Taewon Lee	FDA/NCTR	71	Steve M Clark	GlaxoSmithKline
33	Quan-Zhen Li	Univ Texas Southwestern Medical Center (UTSW)	72	Lisa J. Croner	Biogen Idec
34	Wenyuan Li	Univ. of Texas at Dallas	73	Jing Han	FDA/CBER
35	Yun Lian	Univ Texas Southwestern Medical Center (UTSW)	74	Gene A. Pennello	FDA/CDRH
36	Ying Liu	Univ. of Texas at Dallas	75	Alan H. Roter	Iconix
37	Edward K. Lobenhofer	Cogenics, a Division of Clinical Data, Inc.	76	Tieliu Shi	Chinese Academy of Sciences
38	Tim McDaniel	Illumina	77	Richard Simon	National Cancer Institute
39	Nan Mei	FDA/NCTR	78	Charles Wang	UCLA/Cedars-Sinai
40	Yuri Nikolsky	GeneGo Inc.	79	Sue Jane Wang	FDA/CDER
			80	Liang Zhang	CapitalBio Corporation