The MicroArray Quality Control (MAQC) Project:
An FDA-Led Effort Toward Personalized Medicine

Summary of the 8th MAQC Project Meeting

Development and Validation of Predictive Models Based on Microarray Data

March 24-26, 2008
US FDA, Rockville, Maryland

Summary Author: Leming Shi (leming.shi@fda.hhs.gov)
Summary Date: April 10, 2008
Meeting Organizers: Leming Shi, Jeannette Coleman, Weida Tong, and Federico Goodsaid (FDA)
MAQC Contact: Leming.Shi@fda.hhs.gov, Tel: +1-870-543-7387
MAQC Website: http://edkb.fda.gov/MAQC/

Microarray DAP "for dummies":

"As a consortium, the MAQC should recommend a single consensus Data Analysis Protocol (DAP), simple yet effective, for each microarray platform."

Xijin Ge (South Dakota State University)

The 8th face-to-face MAQC project meeting was held on March 24-26, 2008 at the US Food and Drug Administration's Advisors and Consultants Staff Conference Room in Rockville, Maryland. A total of 101 on-site participants from 60 organizations representing eight countries attended the meeting. In addition, about ten people participated in part or all of the meeting via WebEx or phone. The main objectives of the meeting were: (1) Present Data Analysis Protocols (DAPs) and analysis results; (2) Discuss criteria for selecting MAQC's "candidate" model for each of the 13 endpoints from the six data sets; (3) Discuss a plan for generating additional gene expression and genotyping data; and (4) Discuss manuscript topics, team leaders, and timeline. By March 21, 32 data analysis teams submitted 15,483 models to the MAQC-II. Presentations (PowerPoint or PDF files) and audio recordings of the entire meeting (1.4 GB) are available by contacting Leming.Shi@fda.hhs.gov.

March 24, 2008 (Day One)

Session I-A: MAQC-II Overview and Working Group Updates

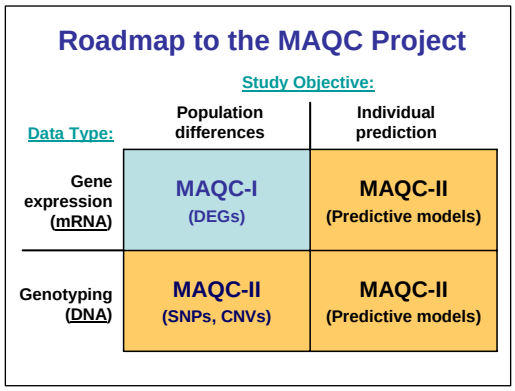
Chair: **Federico Goodsaid** (FDA/CDER)

This Session was aimed at updating MAQC-II participants of the general progress that each WG has made so far and reaffirming each WG's objectives.

- **Robert O'Neil**, Director, Office of Biostatistics, FDA/CDER, welcomed the meeting participants and pointed out the importance of the MAQC-II effort. Dr. O'Neil congratulated to everyone for organizing, managing, and contributing to this unique multi-partner effort, "a model for interaction and synergy". He expected that the outcome of this effort would include good practice and procedures, consensus, consequence and comparisons of different approaches. There are good reasons for a strong regulatory biostatistics interest as considerable methodology is already available to structure the issues and everyone is doing a version of this; it is time to place on sound footing. These

classifiers, best practices, metrics of performance will eventually come to the FDA as part of medical product development and approval. It is therefore important to get things right as early as possible and bring order and good practice to the field. The MAQC-II participants greatly appreciate Dr. O'Neil's advices and encouragement.

- **Leming Shi** (FDA/NCTR) provided an overview of the MAQC project and outlined the agenda for this 8th face-to-face meeting. Leming explained the rationales behind the two phases of the MAQC project. The MAQC-I demonstrated the technical performance of microarray platforms in the identification of differentially expressed genes (DEGs). The objective of MAQC-II is aimed at reaching consensus on the "best practices" for developing and validating predictive models based on microarray data. Reliable and robust predictive models are essential to realize the great promises of personalized medicine. Leming anticipates that a better understanding of the capabilities and limitations of microarray data analysis approaches in clinical and toxicogenomic applications could be reached and recommendations on the development and validation of classifiers may be put forward through the MAQC-II. To accomplish this, we need to explore many different options in analyzing each of the six data sets (a total of 13 endpoints) to generate a unique data set of predictive models through the contributions of many data analysis teams. By the time of the face-to-face meeting, 15,483 models on the 13 endpoints were submitted to the MAQC-II from 32 data analysis teams. The MAQC-II started with microarray gene expression data and has expanded to genotyping data, with the ultimate goal of developing predictive models useful for personalized medicine. Leming emphasized that MAQC is research project; participation is completely voluntary and each participant is expected to cover her/his own costs. He expressed gratitude to the scientific community's enthusiastic participation in and support of the MAQC project. Leming concluded his presentation by reiterating that the MAQC effort is a research collaboration, not a competition, among hundreds of participants. We should work as one team, because we will be judged by the scientific community as a team.



- **Wendell Jones** (Expression Analysis Inc.) gave a overview of the Clinical WG with the team's goals to (1) Understand the behavior of various prediction rules and gene selection methods that may be applied to microarray data sets to generate predictors of clinical outcomes; and (2) Identify and characterize sources of variability in multi-gene prediction results including: a) The impact of tissue acquisition and sample preparation, b) Inter- and intra-laboratory variation in prediction results, and c) Cross-platform performance of prediction results. Wendell outlined a work plan involving the solicitation of data from parties who possess large clinically annotated gene expression data sets that are relevant for the goals of the project. The MAQC members are collectively analyzing the data, compare results, and make recommendations for suitable and/or best practices. New experiments were also suggested to generate data for independent prospective validation and assessment of reproducibility of prediction outcomes. Wendell discussed the administration of the Clinical WG that is also coordinated by Lajos Pusztai (MD Anderson) and Uwe Scherf (FDA/CDRH) and contains more than 200 participating individuals. The three disease areas being analyzed by the MAQC are breast cancer, multiple myeloma, and neuroblastoma. Wendell discussed the formation of a QC subgroup that had assessed the impact of the quality of individual arrays on prediction performance.
- **Richard Judson** (EPA/NCCT) gave an overview of the modeling efforts on the three toxicogenomics data sets: Hamner (mouse lung tumor), Iconix (liver carcinogenicity), and NIEHS (rat necrosis). Richard highlighted some Toxicogenomics-specific issues such as batch effects (time and chemical), small sample sizes, and chemical "domain of applicability". Richard also drew a comparison between

clinical applications and toxicological applications of microarrays. For toxicogenomics, the fundamental question is whether a chemical is toxic or not (in humans). The goal is to predict human outcome using short term test, model species, or tissues. Therefore, the test sample is significantly different from the target being predicted ("Long ago and far away"). Richard cautioned that due to the complexities of toxicogenomics, we should have modest expectations for prediction power of today's models, especially if they are purely driven by statistics with no biological filtering. Richard also briefly introduced the EPA's ToxCast Program, which was designed to prioritize environmental chemicals for further testing based, in part, on genomic profiling of in vitro chemical treatments. Reproducible and validated signatures will be used as part of EPA regulatory prioritization. The Toxicogenomics WG is coordinated by Federico Goodsaid (FDA/CDER) and Richard Judson (EPA/NCCT).

- **Marc Salit** (NIST) discussed what could be learned from the analysis of titration samples. Since the last face-to-face meeting (May 2007), lots of analysis of the pilot titration and recent Agilent titration data were conducted. Walter Liggett and Russ Wolfinger have pursued model-based analysis. Systematic deviation from the model permits identification and characterization of experiment factor effects (e.g. batch effects), with ongoing work focusing on estimating the magnitude of effects. Simon Lin's analysis identified some aberrant behavior in the titration data (still unexplained). There was extensive evaluation of "kinking" genes, but with little statistical evidence to support systematic effects. Because the existing titration experiments with MAQC samples A and B did not include biological variability, the Titration WG is actively planning a new titration experiment to assess relative magnitudes of technical and biological variability. Ron Peterson of Novartis will provide RNA samples from the liver and kidney of 10 normal control rats from an existing study. The liver and kidney samples will be titrated (mixed) according to an experiment design soon to be finalized. The titration samples will be analyzed by multiple platforms (e.g., Affymetrix, Agilent, and Illumina). There is an opportunity to submit the data set for analysis at CAMDA '08 meeting. Several potential manuscripts topics were also discussed. Marc also briefed the MAQC-II of the current status of the External RNA Control Consortium (ERCC); this effort is expected to greatly benefit the gene expression community by providing a panel of well-characterized common external controls. The Titration WG is coordinated by Russ Wolfinger (SAS Institute), Rich Shippy (Affymetrix), and Rick Jensen (Virginia Tech). Marc has been playing a key role in the Titration WG.
- **Greg Campbell** (FDA/CDRH) congratulated the 32 data analysis teams for generating 15,483 models on the 13 endpoints from the six data sets. This creates a serious problem of multiplicity: The performance is overestimated for a reason that is related to the Regression-to-the-Mean effect or the Rookie effect in baseball. The difficulty is that it is unclear how to adjust for this bias. The variance is underestimated but there are multiplicity methods to adjust for this bias. One crude way is to do a Bonferroni adjustment that inflates the variance by the number of classifiers. It is important for each data analysis team to pick a "best" (a better word might be "robust" or "candidate") model for each endpoint. Similarly, the MAQC-II as one team should pick one model for each endpoint before validation data can be distributed. Greg offered some suggestions in picking the candidate model: (1) Parsimonious; (2) Simple preferred over complex; (3) Internal validation performance; (4) Small standard deviation; (5) Range and number of parameters tuned or models built; (6) Well-explained how it was developed; (7) Worry about correlated genes; (8) How did each data analysis team select the one model? A model developed with a biological basis would be preferred. Greg proposed two topics for manuscript development: one based on the principles of predictive model building (based on the SOP document) and one on multiplicity in the project, including the selection process for the candidate model both within each data analysis team as well as within the entire MAQC-II. Greg proposed that RBWG and interested parties plan to meet in the conference room on Monday from 6 pm for about an hour to discuss the criteria for selecting a candidate model for each endpoint. These additional discussions proved to be very helpful and productive. The RBWG is coordinated by Greg Campbell (FDA/CDRH), Lakshmi Vishnuvajjala (FDA/CDRH), and Tim Davison (Asuragen).

- **Federico Goodsaid** (FDA/CDER) presented the working plan for the Genome-Wide Association WG (GWAWG) and drew the synergy between the MAQC effort and the FDA Voluntary eXploratory Data Submission (VXDS) program. Regulatory review of microarray data takes parallel paths: (1) Reconstruct what the sponsor did and apply alternative assumptions in a parallel analysis in order to have a fundamental understanding of data analysis protocols and identify factors affecting variation in classifier results; (2) Develop database, analysis and pathway tools to match specifications of those available to sponsors. The goal of VXDS is biological interpretation. Genome-wide association data sets have been submitted to the VXDS. It is important to know for the Agency how many ways can we NOT match what the sponsor did (e.g., QC, normalization, analysis, and biological interpretation). The FDA has developed collaborations to better understand the analysis of many different GWA data sets. The GWAWG was proposed and discussed at the MAQC May 2007 meeting. Many WebEx seminars were set up for participants to get familiar with the nature and analysis of GWA data. A draft working plan has been discussed during many conference calls. The GWAWG is focusing on the HapMap and Wellcome Trust Case Control Consortium (WTCCC) data sets for which .CEL files are available to MAQC-II participants. The HapMap Data Analysis Team is being coordinated by Huixiao Hong (FDA/NCTR) and the WTCCC Data Analysis Team is being coordinated by Li Zhang (FDA/CDER) and Silvia Vega (Rosetta). Data analysis has started since March 2008. The GWAWG is coordinated by Federico Goodsaid, Huixiao Hong, and Nick Xiao (NCI/SAIC).

Session I-B: **Validation (Blinded) Data Sets**
Chair: **Lajos Pusztai** (MD Anderson Cancer Center)

In this Session, we discussed the current status of the existing validation data sets and a proposal for generating additional gene expression and genotyping data for evaluating the predictive models being developed by the MAQC-II.

- **Lajos Pusztai** (MD Anderson Cancer Center), as a breast cancer clinician and practitioner of gene expression in the clinic, emphasized the needs for demonstrated validation, robustness, and clinical utility of microarray based signatures. For the validation, assay performance should be demonstrated based on independent samples from similar patient population (for treatment, for clinical variables), with the same sampling methodology (biopsy type, fixation/storage), the same analytical technique (measurement platform, SOP), and the same prediction rule (normalization, cut off values). A robust model is expected to work under more stringent situations where assay conditions are variable. Variable conditions could include (1) Slight deviations from SOP; (2) Different measurement platforms (e.g., Affymetrix U133A versus Plus 2.0, Affymetrix versus Agilent, etc.); (3) Different pre-analytical tissue processing (RNAlater versus snap frozen); (4) Different tissue sampling methods (fine needle versus core needle versus surgical biopsies); and (5) Different patient population (similar treatment type, different clinical variables). Even validated and robust assays (signatures) may have limited or no clinical value if the following cannot be demonstrated: (1) Discriminating power (absolute rate of events in prediction groups); (2) Superiority over existing tests; and (3) Improved clinical outcome because of using the test. All clinical trials are a compromise between the ideal design and what is practically feasible! Lajos described the clinical information and demographics of the breast cancer validation set of 100 new cases.
- **Pierre Bushel** (NIH/NIEHS) described the NIEHS validation set of 204 samples. **Benedikt Brors** (DKFZ) described the neuroblastoma validation set of 253 cases (506 arrays from Agilent two-color platform). **Wendell Jones**, on behalf of **Yiming Zhou** of the UAMS, described the MM validation set of 214 new cases. The validation sets for the Hamner (40 samples) and Iconix (201 samples) studies were not presented but have already been submitted to the MAQC-II. Except for the NB study, the validation set and the corresponding training set were generated using the same microarray platforms for the other five studies. For the NB study, 100 probes available in the training set arrays are no

longer available in the validation set. Benedikt Brors will distribute the 100 probes to the data analysis teams so that these probes will be explicitly excluded in the model development.

Table 1. MAQC-II Existing Validation Sets

No.	Date Set Code	Endpoint Code	Number of Validation Set Samples	Availability	Number of Training Set Samples
1	Hamner*	A	40	Immediately	70
2	Iconix	B	201	Immediately	216
3	NIEHS*	C	204	Immediately	214
4	BR*	D, E	100	June 3, 2008	130
5	MM*	F, G, H, I	214	Immediately	340
6	NB*	J, K, L, M	253 (+10)	Immediately	246

*Each data analysis team is required to sign the MAQC Confidential Information Disclosure and Transfer Agreement (CIDTA) with the data provider. E-mail Rusty Thomas (rthomas@thehamner.org, Hamner), Richard Paules (paules@niehs.nih.gov, NIEHS), Lajos Pusztai (lpusztai@mdanderson.org, BR), Sharon Kaufman (sekaufman@uams.edu, MM), and André Oberthuer (andre.oberthuer@uk-koeln.de, NB) to obtain the CIDTA for signature.

- **Leming Shi** (FDA/NCTR) presented a proposal for generating additional gene expression and genotyping data for evaluating the performance of predictive models being developed by the MAQC-II. This proposal was largely based on the availability of samples and intended to demonstrate the reproducibility of model prediction when the same set of samples are analyzed in different microarray laboratories. The breast cancer prediction reproducibility study was first proposed by Dr. W. Fraser Symmans of the MD Anderson Cancer Center during the 6th MAQC face-to-face meeting, November 2006 for testing existing breast cancer signatures. In total, 975 Agilent microarrays and 825 Affymetrix gene expression microarrays are needed (a total of 1,800 gene expression microarrays). In addition, 2,000 Affymetrix SNP 6.0 microarrays are needed to generate a unique, integrative data set with SNP, CNV, and gene expression data for the same set of multiple myeloma patients. Please contact Leming Shi if you may be able to contribute to this effort.

Table 2. MAQC-II Needs for Microarrays to Generate Additional Validation Data Sets

Disease/Study	Number of Samples	Number of Test Sites	Number of Microarrays	Microarray Platform	Sample Sources
Hamner	100	1	100	Affymetrix Mouse 430 2.0	Hamner
NB (Neuroblastoma)	350	1	350	Agilent One-Color	Univ. of Cologne
MM (Multiple Myeloma)	100	1	100	Affymetrix U133Plus2	UAMS
	2,000 (DNA)	*1*	*2,000*	*Affymetrix SNP 6.0*	
BR (Breast Cancer)	125	3+1+1	625	Affymetrix U133A and Plus2	MDACC etc.
	125	3+1+1	625	Agilent One- and Two-Color	

Sessions I-C to II-C (Day One and Day Two): Data Analysis Protocols (DAPs) and Analysis Results

Sessions I-C to II-C were organized for each data analysis team to present its Data Analysis Protocol (DAP) and the corresponding data analysis results. Among the 35 data analysis teams, 32 teams submitted models to the MAQC-II by March 21, 2008, and 25 teams presented their DAPs and analysis results at the meeting. These teams' efforts in model development were crucial contributions to the MAQC-II and formed the basis of much of the presentations and discussions at the meeting. We are grateful to their dedications to the MAQC-II project.

Table 3. Presentations on Data Analysis Protocols (DAPs) and Analysis Results

Session I-C Chair: **Greg Campbell** (CDRH/FDA)	
1. CAS (Chinese Academy of Sciences, China)	**Tieliu Shi**
2. CDRH (Center for Devices and Radiological Health, FDA)	**Samir Lababidi**
3. CIPF (Centro de Investigacion Principe Felipe, Spain)	**Ignacio Medina**
4. Cornell (Weill Medical College of Cornell University)	**Fabien Campagne**
5. DKFZ (German Cancer Research Center, Germany)	**Benedikt Brors**
6. EPA (U.S. Environmental Protection Agency)	**Fathi Elloumi and Zhen Li**
Session I-D Chair: **Wendell Jones** (Expression Analysis)	
7. GeneGo (GeneGo Inc.)	**Andrej Bugrim**
8. UIUC (University of Illinois at Urbana-Champaign)	**Yanen Li**
9. NCTR (National Center for Toxicological Research, FDA)	**Huixiao Hong**
10. NIEHS (National Institute of Environmental Health Sciences)	**Jeff Chou and Jianying Li**
Session II-A Chair: **Lakshmi Vishnuvajjala** (CDRH/FDA)	
11. NWU (Northwestern University)	**Simon Lin**
12. Princeton (Princeton University)	**Yichao Wu**
13. SAI (Systems Analytics Inc.)	**John Zhang**
14. SAS (SAS Institute Inc.)	**Russ Wolfinger**
15. SIB (Swiss Institute of Bioinformatics, Switzerland)	**Vlad Popovici**
16. Spheromics (Spheromics, Finland)	**Max Bylesjö** (Umeå Univ.)
Session II-B Chair: **Tim Davison** (Asuragen)	
17. SuperArray (SuperArray Bioscience Corporation)	**Guozhen Liu**
18. Tsinghua (Tsinghua University, China)	**Shicai Fan**
19. USM (University of Southern Mississippi)	**Venkata Thodima**
Session II-C Chair: **Kenneth Hess** (MD Anderson Cancer Center)	
20. JHSPH (Johns Hopkins Bloomberg School of Public Health)	**Rafael Irizarry**
21. GT (Georgia Institute of Technology – Emory University)	**May Wang**
22. SDSU (South Dakota State University)	**Xijin Ge**
23. KU (University of Kansas)	**Luke Huan**
24. ABT (Abbott Laboratories)	**Viswanath Devanarayan**
25. UAMS (University of Arkansas for Medical Sciences)	**Yiming Zhou**
The following seven teams submitted models to the MAQC-II but were unable to present their DAPs and results at the meeting: 1. Almac (Almac Diagnostics, UK, **Juergen von Frese**); 2. CBC (CapitalBio Corporation, China, **Liang Zhang**); 3. FBK (Fondazione Bruno Kessler, Italy, **Cesare Furlanello**); 4. Ligand (Ligand Pharmaceuticals, **Wen Luo**); 5. Roche (Roche Palo Alto LLC, **Mark Fielden**); 6. UCLA (Cedars-Sinai Medical Center of UCLA, **Xutao Deng**); and 7. ZJU (Zhejiang University, China, **Xiaohui Fan**).	

Before individual data analysis teams' presentations, **Russ Wolfinger** (SAS Institute) and **Kenneth Hess** (MD Anderson Cancer Center) presented an overview of the 15,483 models submitted to the MAQC-II by March 21, 2008. Their meta-analyses clearly showed the diversity of methodologies used in model development and performance assessment. It was clear that among the 13 endpoints, some are much easier to predict than others, and some are extremely hard to predict. For some endpoints, the performance estimates of the "candidate" models from various teams showed dramatic differences, indicating overfitting might have been a problem in some DAPs. **Wendell Jones** (Expression Analysis) and **John Zhang** (Systems Analytics) also provided their meta-analysis results on the submitted models. Each data analysis team's presentation was followed by questions/answers and constructive discussions. MAQC-II participants are encouraged to review these presentations to better appreciate each team's effort.

After the originally scheduled sessions on March 24 were completed, the RBWG organized a follow-up discussion session for about 1 hour on the criteria and procedure for selecting MAQC's "candidate" models for validation. We agreed to use the term "candidate", instead of "best", to represent the model to be selected for "validation" purpose in order to address multiplicity issues. It was also agreed that each selected "candidate" model should be evaluated/reconstructed by at least one independent group to ensure that the process and the results are reproducible. The need for some data analysis teams to make corrections to their DAPs and rerun the calculations became obvious.

Individual data analysis teams' presentations continued on March 25, accompanied with many good discussions following each presentation. Further discussions on the selection of the MAQC's "candidate" models were coordinated by **Bob Wagner** (FDA/CDRH) and **Wendell Jones** (Expression Analysis). The following consensus was reached:

1. Analysis teams will have the opportunity to refine their model-building efforts in compliance with SOP especially regarding the methods that address assessment of predictive performance (incl. precision). Address potential issues that referees may raise; Feature selection embedded within internal/cross validation; Global normalization methods such as quantile and RMA frowned on unless using a reference set. In general, data leakage is to be reduced or eliminated. Decision will come soon as this effect may be published; Batch effect removal? 5-fold CV is strongly recommended for comparison purposes only to assist the Steering Committee in selecting the candidate model per endpoint. However, the Steering Committee should clearly specify exactly how the 5-fold CV should be carried out and how many replicates ($\geq$10) and how estimates of variance will be calculated.
2. By consensus, the Steering Committee will ultimately choose the candidate model per endpoint, selecting from a pool recommended by the RBWG. Once a candidate model is being seriously considered, at least one group performing related work will be asked to duplicate the model and model performance results (not exactly, but within statistical reasonableness). Portability of your model and parameter selection methods will be important. Creating a portable script is good. Describing from a more detailed DAP and creating independently is also OK, sometimes preferred. The Steering Committee will be interested in models that address clinical covariates in addition to strictly genomic-based models.
3. The MAQC would still ask groups to rank their models by endpoint in terms of the ones they feel have the most desirable properties. Related to this, the MAQC asks that groups report a more complete set of models that were considered. The MAQC leadership will come back with recommendations for standardizing entries and reporting levels (i.e., certain models that only change certain parameters learned during training would not be reported).
4. Each model, "candidate" or not, will be used to predict the External validation data sets.

Leming Shi reiterated the main objectives of the MAQC-II data analysis efforts: establishing "baseline/good/best" practices (procedures) applicable to future microarray data sets in developing and validating microarray-based predictive models. Therefore, it is imperative to understand which Modeling

Factors are more important in determining the Interval and External validation performances of a predictive model, and why do some Data Analysis Protocols (DAPs) succeed or fail in external prediction. This could be accomplished by creating and mining a large data set of many *predictive models* developed by the data analysis teams. Each row represents a model, which can be characterized by (1) a set of Modeling Factors; (2) a set of performance metrics in Internal Validation; and (3)) a set of performance metrics in External Validation (Figure 1). It was agreed that the parameters requested in the model submission template are sufficient, but some standardization is required so that the result data set of models could be mined. The data set of 15,483 models submitted as of March 21, 2008 did not appear to be ideal because the number of models submitted by each data analysis team is dramatically different (8,580 versus 1!) and could be dominated by a few teams that submitted the most models. To help create a "dream" data set of predictive models, each data analysis team is urged to do the following:

1. Apply the same DAP to all 13 endpoints from the six data sets.
2. Submit to the MAQC-II ***all*** models that have been explored by the team. "Larger sample size is better"; it also help the RBWG assess the severity of multiplicity. Remember that "bad" models are equally important as the "best/candidate" models for the MAQC-II to develop good practices.
3. Use the same internal validation procedure to assess a model's performance – it has been decided that a 10x stratified 5-fold cross-validation should be used by all teams (see **Russ Wolfinger**'s e-mail on March 27, 2008 to the MAQC mailing list). A few teams may choose to provide extra internal validation performance estimates using other procedures that will be compared with 10x 5F-CV.
4. Each model should be accompanied by the same six performance metrics: MCC, Accuracy, Sensitivity, Specificity, AUC, and RMSE. See Russ' e-mail for instructions to calculate RMSE.

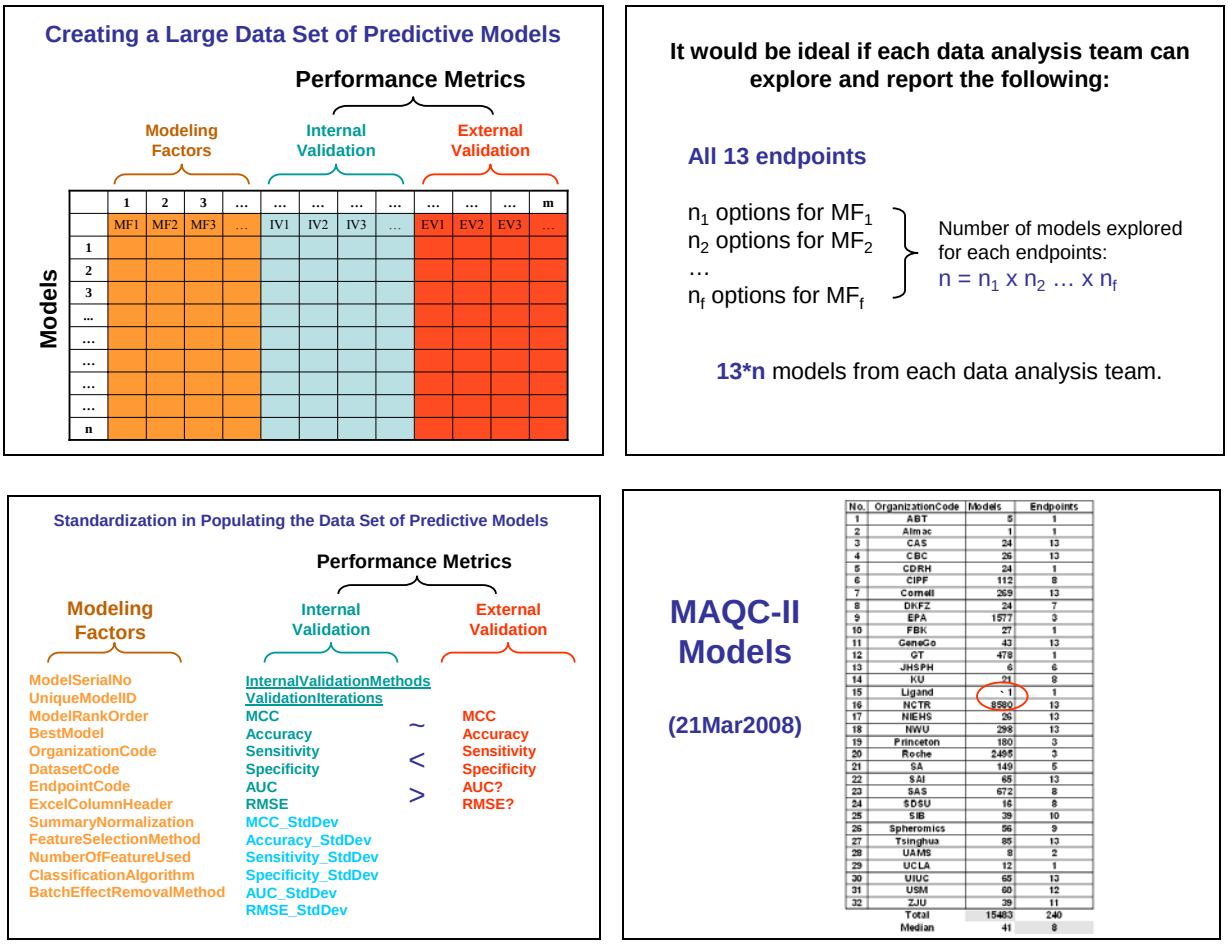


Figure 1. MAQC-II data analysis: Generating a large data set of predictive models for meta-analysis

Data analysis teams were urged to follow the RBWG SOP to avoid obvious problems such as gene selection bias. We also had lengthy discussions on multiple-array based normalization such as RMA. It was generally agreed upon that a normalization method should be applicable to one single new sample to be predicted. refRMA is one of such method. **Wendell Jones** prepared a template for reporting model prediction results on the validation sets, but we did not have time to discuss it. This will be discussed over TC or WebEx as we get closer to the distribution of the validation data sets.

March 26, 2008 (Day Three)

Sessions III-A and III-B: Manuscript Preparation

Chairs: **Jim Fuscoe** and **Leming Shi** (FDA/NCTR)

In these two Sessions, meeting participants made short presentations to outline the many topics around which manuscripts might be developed. The presentations and follow-up discussions were not meant to make final decisions on a set of manuscripts to be developed under the MAQC-II. Rather, they served as an opportunity for exchanging creative ideas to guide the next phases of data analyses, keeping in mind what needs to be done for a topic to be developed into a solid manuscript. Many proposals were made during the meeting or immediately following the meeting (Table 4). Additional manuscript topics may be proposed as more data analyses are to be conducted and new ideas emerge. It is expected that proposals with similar objectives will eventually be merged. If you are interested in contributing to a manuscript topic, please contact the coordinator. The coordinators have been advised to be as inclusive as possible.

Session III-C: Genome-Wide Association Working Group (GWAWG)

Chair: **Federico Goodsaid** (FDA/CDER)

This Session was aimed at updating MAQC-II participants of what the GWAWG has been doing or is planning to do. Federico gave an overview of the WG and the agenda for this session.

- **Nick Xiao** (NCI/SAIC) provided an overview of the genotyping data sets that could be available for the MAQC-II to analyze.
- **Huixiao Hong** (FDA/NCTR) presented a data analysis plan focusing on how different genotype calling algorithms and QC processes impact the outcome in terms of differentiating SNP lists. HapMap data will be a focus. To join this effort, contact huixiao.hong@fda.hhs.gov.
- **Li Zhang** (FDA/CDER) presented a plan on the analysis of the WTCCC data set, starting with the coronary artery disease (CAD) samples. To join this effort, contact li.zhang@fda.hhs.gov.
- **Christophe Lambert** (Golden Helix) talked about genome-wide copy number variation (CNV) association and batch effects observed in the WTCCC data sets. Christophe also built disease predictive models with CNV.
- **Russ Wolfinger** (SAS Institute) gave a live demonstration on the analysis of genotyping data with JMP Genomics focusing on CNV in the coronary artery disease (CAD) samples of the WTCCC. Russ pointed out that the GWAWG should make prediction modeling the ultimate goal of data analysis and advocated the direct use of intensity values in analysis.
- The CNV data analysis team was formed.

Concurrent Session III-C: Titration Working Group

Chair: **Marc Salit** (NIST)

The Titration Working Group's discussions focused on the design of a new titration experiment aimed at addressing biological variability in tissue-mixing. The experimental design has been finalized soon after the face-to-face meeting. RNA samples have been mixed by **Ron Peterson** (Novartis) and Affymetrix arrays have been processed. The RNA samples have been shipped to Agilent and Illumina for processing.

Table 4. Proposed Manuscript Topics

No.	Proposed Manuscript Topic	Coordinator	E-mail
1	The "Main" Manuscript: MAQC-II Overview and "Good" Practices	Leming Shi (NCTR/FDA)	leming.shi@fda.hhs.gov
2	Prediction Reproducibility of Breast Cancer Signatures	W. Fraser Symmans (MDACC)	fsymmans@mdanderson.org
3	Cross-Platform Transferability (NIEHS Data)	Weida Tong NCTR)	weida.tong@fda.hhs.gov
4	Cross-Tissue Prediction (NIEHS Data)	Pierre Bushel (NIEHS)	bushel@niehs.nih.gov
5	Array Data Quality	Wendell Jones (EA)	wjones@expressionanalysis.com
6	Normalization Methods	Kenneth Hess (MDACC)	khess@mdanderson.org
7	Batch Effects	John Zhang (SAI)	johnz@systemsanalytics.com
8	Cross-Batch/Platform Prediction	Andrej Bugrim (GeneGo)	andrej@genego.com
9	"Statistical Methodologies" (RBWG SOP)	Greg Campbell (CDRH)	greg.campbell@fda.hhs.gov
10	Multiple Titration Manuscripts	Marc Salit (NIST)	salit@nist.gov
11	GWAWG: Genotype Calling	Huixiao Hong (NCTR)	huixiao.hong@fda.hhs.gov
12	One-color *vs*. Two-color: Neuroblastoma (Agilent Platform)	Benedikt Brors (DKFZ) Russ Wolfinger (SAS)	b.brors@dkfz-heidelberg.de russ.wolfinger@sas.com
13	Multiple Myeloma Manuscript	Yiming Zhou (UAMS)	yzhou@uams.edu
14	Uncertainties in the Classifier Problem	Weijie Chen (CDRH)	weijie.chen@fda.hhs.gov
15	Comparative Analysis of Modeling Practices	Weida Tong (NCTR)	weida.tong@fda.hhs.gov
16	Multi-Path Learning	Andrej Bugrim (GeneGo)	andrej@genego.com
17	Factors Affected Toxicity Prediction	Richard Judson (EPA)	judson.richard@epa.gov
18	Candidate Models	Wendell Jones (EA)	wjones@expressionanalysis.com
19	Meta-analysis of Gene Features	Youping Deng (USM)	youping.deng@usm.edu
20	Stability of Genomic Signatures	Cesare Furlanello (FBK)	furlan@fbk.eu
21	Gene Selection Methods	Simon Lin (NWU)	s-lin2@northwestern.edu
22	Breast Cancer Manuscript	Lajos Pusztai (MDACC)	lpusztai@mdanderson.org
23	Good Clinical Practices (GCP) Document	Guy Tillinghast (Riverside)	guy.tillinghast@rivhs.com
24	Multiplicity Issues in the MAQC-II	Greg Campbell (CDRH)	greg.campbell@fda.hhs.gov
25	MAQC, VXDS, and FDA Guidance	Federico Goodsaid (CDER)	federico.goodsaid@fda.hhs.gov
	……		

MAQC-II To Do List and Tentative Timeline

1. Meeting summary (Shi) – April 10, 2008
2. Finalize "Model Summary" template with detailed instructions (Wolfinger) – April 14, 2008
3. Re-run all analyses with 10x 5F-CV (Data Analysis Teams - DATs) – Now
4. Sign confidential agreements with data providers (Table 1) – Now
5. Submit models (DATs) – **by May 5, 2008**
6. Distribute the data set of models (Shi) – May 7, 2008
7. Select "candidate" models (RBWG and Steering Committee) – May 15, 2008
8. Distribute validation data sets (Shi) – May 16, 2008
9. Submit prediction results (DATs) – June 2, 2008
10. Swap training and validation sets (DATs)
11. Prepare manuscripts (All)
    - VO: April 28 (Detailed manuscript outlines)
    - V1: July 14 (Draft manuscript)
    - V2: Aug. 4 (Revised)
    - V3: Aug. 18 (Revised, ready for institutional clearance)
    - V4: Sept. 1 (Revised, almost ready for peer review)
    - VS: Sept. 8, 2008 (Submission for peer review)

Table 5. Participants of the 8th MAQC Project Meeting, March 24-26, 2008, Rockville, Maryland, USA

No.	Name	Organization	No.	Name	Organization
1	Nicholas Beckloff	Information Management Consultants	53	Yunqing Ma	Panomics
2	Anne Bergstrom Lucas	Agilent	54	Teri Manolio	NIH/NHGRI
3	Vincent Bertholet	Eppendorf Array Technologies	55	Francisco Martinez-Murillo	FDA/CDRH
4	Benedikt Brors	German Cancer Research Center (DKFZ)	56	Matt McCall	Johns Hopkins University
5	Andrej Bugrim	GeneGo Inc.	57	Ignacio Medina	Centro de Investigacion Principe Felipe
6	Pierre Bushel	NIH/NIEHS	58	Richard Moffitt	Georgia Institute of Technology
7	Max Bylesjo	Umeå University	59	Padraic Neville	SAS Institute
8	Fabien Campagne	Weill Medical College of Cornell University	60	Oluwole Odujinrin	Customized Therapeutics LLC
9	Gregory Campbell	FDA/CDRH	61	Robert T O'Neill	FDA/CDER
10	Jennifer G. Catalano	FDA/CBER	62	R. Mitchell Parry	Georgia Institute of Technology
11	Yu-Ling Chang	FDA/CDRH	63	Roger G. Perkins	FDA/NCTR (ICF International)
12	Weijie Chen	FDA/CDRH	64	Ron Peterson	Novartis
13	Jeff Chou	NIH/NIEHS	65	John Phan	Georgia Institute of Technology
14	Jeannette F. Coleman	FDA/NCTR	66	Vlad Popovici	Swiss Institute of Bioinformatics
15	Timothy S. Davison	Asuragen	67	Lajos Pusztai	MD Anderson Cancer Center
16	Arkendra De	FDA/CDRH	68	Jacques D. Retief	Illumina
17	Francoise de Longueville	Eppendorf Array Technologies	69	Marc Salit	NIST
18	Francesca Demichelis	Weill Medical College of Cornell University	70	Andreas Scherer	Spheromics
19	Xutao Deng	UCLA/Cedars-Sinai	71	Martin Schumacher	Novartis
20	Youping Deng	University of Southern Mississippi	72	Joe Shambaugh	Genedata (USA) Inc.
21	Viswanath Devanarayan	Abbott	73	Leming Shi	FDA/NCTR
22	Pan Du	Northwestern University	74	Tieliu Shi	Chinese Academy of Sciences
23	Fathi Elloumi	EPA	75	Richard Shippy	Affymetrix
24	Shicai Fan	Tsinghua University	76	Todd H Stokes	Georgia Institute of Technology
25	Yang Feng	Princeton Univeristy	77	Rong Tang	FDA/CDRH
26	Elvene Fong	SuperArray	78	Zivana Tezak	FDA/CDRH
27	James C. Fuscoe	FDA/NCTR	79	Danielle Thierry-Mieg	NIH/NCBI
28	Weiniu Gan	NIH/NHLBI	80	Jean Thierry-Mieg	NIH/NCBI
29	Xijin Ge	South Dakota State University	81	Venkata Thodima	University of Southern Mississippi
30	Federico M. Goodsaid	FDA/CDER	82	Guy Tillinghast	Riverside Cancer Care Center
31	Lei Guo	FDA/NCTR	83	Ram C. Tiwari	NIH/NCI
32	Peter Herzer	Eppendorf Biochip Systems	84	Lakshmi Vishnuvajjala	FDA/CDRH
33	Kenneth Hess	MD Anderson Cancer Center	85	Robert F Wagner	FDA/CDRH
34	Huixiao Hong	FDA/NCTR	86	Stephen J. Walker	Wake Forest University
35	Luke (Jun) Huan	University of Kansas	87	May Dongmei Wang	Georgia Institute of Technology
36	Nina L. Hunter	FDA/CDRH	88	Sue Jane Wang	FDA/CDER
37	Rafael A. Irizarry	Johns Hopkins University	89	Wei Wang	Cornell University
38	Roderick V. Jensen	Virginia Bioinformatics Institute	90	Katrin Welzel	Eppendorf Biochip Systems GmbH
39	Wendell D. Jones	Expression Analysis	91	Russell D Wolfinger	SAS Institute
40	Jungnam Joo	NIH/NHLBI	92	Yichao Wu	Princeton Univeristy
41	Richard Judson	EPA	93	Yan Xiang	SuperArray
42	Mike Kuziora	Gene Logic	94	Chunlin Xiao	Applied Biosystems
43	Samir Lababidi	FDA/CDRH	95	Nianqing Xiao	NIH/NCI (SAIC)
44	Christophe Lambert	Golden Helix	96	Qian Xie	Information Management Consultants, Inc.
45	Nick Lazaridis	Gene Express	97	Jingping Yang	SuperArray
46	Jianying Li	NIH/NIEHS	98	John Zhang	Systems Analytics
47	Yanen Li	University of Illinois at Urbana-Champaign	99	Li Zhang	FDA/CDER
48	Zhen Li	EPA	100	Xiaobo Zhou	The Methodist Hospital Research Institute
49	Walter Liggett	NIST	101	Yiming Zhou	University of Arkansas for Medical Sciences
50	Simon Lin	Northwestern University	The 101 meeting participants came from 60 organizations from eight countries. In addition, about tem people participated in all or part of the three-day meeting via WebEx or phone.		
51	Guozhen (Gordon) Liu	SuperArray			
52	Jean Lozach	Illumina			

Please review the great suggestions and comments from Dr. Xijin Ge (South Dakota State University) and Dr. Martin Schumacher (Novartis Pharma AG). We'll discuss these during upcoming conference calls. –Leming Shi

Suggestions from Dr. Xijin Ge
South Dakota State University
xijin.ge@sdstate.edu
March 27, 2008

1. I propose that data analysis teams report their predictions for each sample by a number within the interval [-1,+1]. A number of -1 will indicate 100% sure that this sample belongs to the negative class, while +1 will mean 100% sure that it will be positive class. Zero means "no call". Of course, if a team still wants to (or have to per their algorithm) report binary results, it could use -1 or 1 only. For each dataset, data analysis teams might also need to provide their cutoff value for making confident prediction.
2. Classifiers with extremely small number of genes (<10) suffer from serious robustness issue. As highlighted in my presentation (and I shamelessly give the link to our paper here: http://hc.ims.u-tokyo.ac.jp/JSBi/journal/GIW03/GIW03F004/index.html), up to 80% of the positive calls could be false positives when tested with a large number of samples that does not belong to neither positive nor negative class. Furthermore, as a diagnostic tool, there's no point in using microarrays when such small number of genes are involved. PCR assays or other methods could be more accurate without significant cost increase.
3. Many teams simply rank the average MCC or other metrics and select their best model. It is not surprising that we ended up with choosing different algorithms at different endpoints. We might use a little bit of statistics in evaluating our models by asking ourselves: Is the performance of the best model significantly different from that of the second best model? We could simple do a student T test to compare the 10 MCC's we got in the best model with the 10 MCC's from the second best model. The same could be done between the top model and the third one, fourth one etc. My guess is there's no significant difference between many of our models.
4. If a team works with many classification algorithms, please treat each algorithm fairly. Understand all algorithms and configure them so that all work best. I don't know about other people, but I myself have my own favorite algorithm and I find myself unconsciously spending more time in tuning up my favorite one. For example, in KNN if we fix K at 1, that will not perform very well.

Suggestions from Dr. Martin M. Schumacher
Novartis Pharma AG
martin.schumacher@novartis.com
April 8, 2008

Dear Leming,

Greetings from Switzerland!

I really enjoyed the meeting in Rockville. Thank you very much for giving me the opportunity to attend.
The meeting was very well organized and I want to thank you for all your efforts.

As discussed after the meeting I want to share some of my impressions and concerns with you. As I am a newcomer to MAQC-II please bear with me if I am addressing issues which were already discussed (and solved) in the past. I will make most statement in a short telegraph style.

In general I believe that the MAQC-II is a very important activity and that we should make every effort to make it as successful as possible. However, I'm afraid that several important parts do not get the emphasis they deserve. In short, I believe that out of the three steps shown below
- Assessment, preprocessing and transformation of the raw gene expression data
- Building of a classifier
- Assessment of the suitability of the test samples for class prediction before doing the actual class prediction

The first and the last step don't get the attention they deserve. I believe that the experts in the field of predictive genomics would agree that if
- the training & test data are of good quality
- the different classes are sufficiently populated & the overall sample size is big enough
- the data is information rich (eg big fold-changes of some genes between the classes)
- the class label is reliable/reproducible
- the test samples come from the same population as the training data

any classifier (even very simple ones) would do a good job. The implication of this statement is that we have to make sure that all the assumption stated above hold or are taken care of.

Please find below a series of thoughts in a non-specific order.

- For all training data sets a post-hoc sample size estimation should be made by the different analysis teams after their specific data pre-treatments. I propose the approach suggested recently by Dobbin et al (Clin Cancer Res 2008, 14(1):108). Also the number of features needed for the classifier could be estimated a-priori. If the sample size is not big enough, the data set should not be used (or labeled as lower quality).
- Also for the test set a sample size estimation should be done using, for example, a pre-defined lower limit of a 1-sided confidence interval.
- It should be verified that the methods used for batch-effect removal don't make use of batch labels and can work with prospective data
- The quantitative level of reproducibility of the class labels should be assessed and stated. If this number is for example 0.8 (and the reference method is believed to be true), then this number would be the maximum obtainable performance of a classifier (and not 1). Classifiers with bigger performance measures above this number are at risk of being overfitted.
- The suitability of the test data for class prediction should be assessed. The clinical population or some technical procedures might have changed (eg new comedications) introducing a bias in the data. A simple projection method like PCA using the same features as the classifier could provide very useful information. A simple chart with Hotelling's T^2 (or the Mahalanobis distance) and the squared projection error (residual) [and the corresponding critical limits] of all test samples based on an appropriate PCA model (build with the training data) would assess the matching of training and test data. If some (or all) test samples do not pass the test, their class label should not be predicted.

It is my hope that my input is of value for the MAQC-II project. Please feel free to distribute to other colleagues. I am willing to discuss any of the points in this mail.
I look forward to hearing from you.

Kind regards,
Martin

FINAL AGENDA
(Updated 21MAR2008)

The MicroArray Quality Control (MAQC) Project:

An FDA-Led Effort Toward Personalized Medicine

The 8th MAQC Project Meeting

Development and Validation of Predictive Models

Monday-Wednesday
March 24–26, 2008
9:00 am – 6:00 pm Eastern Daylight Time

US FDA
Advisors and Consultants Staff Conference Room
Room 1066
5630 Fishers Lane
Rockville, MD 20857, USA

Meeting Objectives:

1. Present Data Analysis Protocols (DAPs) and analysis results;
2. Select MAQC's "best" model for each of the 13 endpoints from the six data sets;
3. Finalize a plan for generating additional gene expression and genotyping data;
4. Decide on manuscript topics, team leaders, and timeline.

Leming.Shi@fda.hhs.gov

Tel: +1-870-543-7387

http://edkb.fda.gov/MAQC/

Participants should consider information exchanged during the MAQC meeting as confidential.

Monday, March 24, 2008 (Day One)

8:00 am	Registration & Continental Breakfast	
Session I-A: MAQC-II Overview and Working Group Updates Chair: Federico Goodsaid (CDER/FDA)		
9:00 am	Welcoming Remarks	Robert O'Neill (CDER/FDA)
9:20 am	Overview of MAQC-II and Meeting Agenda	Leming Shi (NCTR/FDA)
9:45 am	Clinical Working Group	Wendell Jones (Expression Analysis)
9:55 am	Toxicogenomics Working Group	Richard Judson (EPA)
10:05 am	Titration Working Group (and ERCC Update)	Marc Salit (NIST)
10:20 am	Regulatory Biostatistics Working Group (RBWG)	Greg Campbell (CDRH/FDA)
10:30 am	Genome-Wide Association Working Group (GWAWG)	Federico Goodsaid (CDER/FDA))
10:40 am	Discussion	All
11:00 am	Coffee Break & Poster View	
Session I-B: Validation (Blinded) Data Sets Chair: **Lajos Pusztai** (MD Anderson Cancer Center)		
11:30 am	Existing Validation Sets: Hamner (Mouse Lung Tumor) Iconix (Rat Liver Carcinogen) NIEHS (Rat Necrosis) BR (Breast Cancer) MM (Multiple Myeloma) NB (Neuoblastoma)	Lajos Pusztai Pierre Bushel (NIEHS/NIH) Lajos Pusztai (MDACC) Yiming Zhou (UAMS) Benedikt Brors (DKFZ)
12:00 pm	Generating Additional Gene Expression and Genotyping Data	Leming Shi
12:05 pm	Contributions Are Needed from MAQC-II Participants	Manufacturers, Service Providers, and All
12:30 pm	Lunch & Poster View	
Session I-C: Data Analysis Protocols (DAPs) and Analysis Results (1) Chair: **Greg Campbell** (CDRH/FDA)		
2:00 pm	1. CAS (Chinese Academy of Sciences, China)	Tieliu Shi
2:20 pm	2. CDRH (Center for Devices and Radiological Health, FDA)	Samir Lababidi
2:40 pm	3. CIPF (Centro de Investigacion Principe Felipe, Spain)	Ignacio Medina
3:00 pm	4. Cornell (Weill Medical College of Cornell University)	Fabien Campagne
3:20 pm	5. DKFZ (German Cancer Research Center, Germany)	Benedikt Brors
3:40 pm	6. EPA (U.S. Environmental Protection Agency)	Zhen Li and Fathi Elloumi
4:00 pm	Coffee Break & Poster View	
Session I-D: Data Analysis Protocols (DAPs) and Analysis Results (2) Chair: **Wendell Jones** (Expression Analysis)		
4:30 pm	7. GeneGo (GeneGo Inc.)	Andrej Bugrim
4:50 pm	8. UIUC (University of Illinois at Urbana-Champaign)	Yanen Li
5:10 pm	9 NCTR (National Center for Toxicological Research, FDA)	Huixiao Hong
5:30 pm	10. NIEHS (National Institute of Environmental Health Sciences)	Jianying Li and Jeff Chou
5:50 pm	Discussion	All
6:00 pm	Adjourn Day One	

Tuesday, March 25, 2008 (Day Two)

8:00 am	Continental Breakfast	
Session II-A: Data Analysis Protocols (DAPs) and Analysis Results (3) Chair: **Lakshmi Vishnuvajjala** (CDRH/FDA)		
9:00 am	11. NWU (Northwestern University)	Simon Lin
9:20 am	12. Princeton (Princeton University)	Yichao Wu
9:40 am	13. SAI (Systems Analytics Inc.)	John Zhang
10:00 am	14. SAS (SAS Institute Inc.)	Russ Wolfinger
10:20 am	15. SIB (Swiss Institute of Bioinformatics, Switzerland)	Vlad Popovici
10:40 am	16. Spheromics (Spheromics, Finland)	Max Bylesjö (Umeå Univ.)
11:00 am	Coffee Break & Poster View	
Session II-B: Data Analysis Protocols (DAPs) and Analysis Results (4) Chair: **Tim Davison** (Asuragen)		
11:30 am	17. SuperArray (SuperArray Bioscience Corporation)	Guozhen Liu
11:50 am	18. Tsinghua (Tsinghua University, China)	Shicai Fan
12:10 pm	19. USM (University of Southern Mississippi)	Youping Deng
12:30 pm	Lunch & Poster View	
Session II-C: Data Analysis Protocols (DAPs) and Analysis Results (5) Chair: **Kenneth Hess** (MD Anderson Cancer Center)		
2:00 pm	20. JHSPH (Johns Hopkins Bloomberg School of Public Health)	Rafael Irizarry
	New Data Analysis Teams:	
2:30 pm	21. GT (Georgia Institute of Technology – Emory University)	May Wang
2:45 pm	22. SDSU (South Dakota State University)	Xijin Ge
3:00 pm	23. KU (University of Kansas)	Luke Huan
3:15 pm	24. ABT (Abbott Laboratories)	Viswanath Devanarayan
	25. Cornell2 (Cornell University)	*Wei Wang*
	26. UAMS (University of Arkansas for Medical Sciences)	*Yiming Zhou*
3:30 pm	*The following teams are not attending, but may be available over the phone to answer questions about their data analysis results.*	*Team Leader*
	1. Almac (Almac Diagnostics, UK)	*Juergen von Frese*
	2. CBC (CapitalBio Corporation, China)	*Liang Zhang*
	3. FBK (Fondazione Bruno Kessler, Italy)	*Cesare Furlanello*
	4. Ligand (Ligand Pharmaceuticals)	*Wen Luo*
	5. NIEHS2 (National Institute of Environmental Health Sciences)	*Jennifer Fostel*
	6. Roche (Roche Palo Alto LLC)	*Mark Fielden*
	7. UCLA (Cedars-Sinai Medical Center of UCLA)	*Xutao Deng*
	8. UML (University of Massachusetts Lowell)	*Dalila Megherbi*
	9. ZJU (Zhejiang University, China)	*Xiaohui Fan*
3:40 pm	Discussion	All
4:00 pm	Coffee Break & Poster View	
Session II-D: Selection of MAQC's "Best" Models and Distribution of Validation Data Sets Co-Chairs: **Bob Wagner** (CDRH/FDA) and **Wendell Jones** (Expression Analysis)		
4:30 pm	Discussion on the Selection of MAQC's "Best" Model for Each Data Set (Endpoint)	All
5:45 pm	Logistics on Distributing Validation Data Sets	Leming Shi (NCTR/FDA)
5:50 pm	Template for Reporting Prediction Results	Wendell Jones (EA)
6:00 pm	Adjourn Day Two	

Wednesday, March 26, 2008 (Day Three)

8:00 am	Continental Breakfast	
Session III-A: Manuscript Preparation (1) Chair: **Jim Fuscoe** (NCTR/FDA)		
9:00 am	1. The "Main" Manuscript	Leming Shi (NCTR/FDA)
About 5 to 10 minutes for each topic.	2. Cross-Site Prediction Reproducibility (Breast Cancer)	Fraser Symmans (MDACC)
	3. Cross-Platform Transferability of Models (NIEHS Data)	Huixiao Hong (NCTR/FDA)
	4. Cross-Tissue Prediction (NIEHS Data)	Pierre Bushel (NIEHS, WebEx)
	5. Array Data Quality and Model Prediction Performance	Wendell Jones (EA)
	6. Normalization Methods and Model Prediction Performance	Rafael Irizarry (JHSPH)
	7. Batch Effect Removal and Model Prediction Performance	John Zhang (SAI)
	8. Cross-Batch/Platform Prediction	Andrej Bugrim (GeneGo)
	9. RBWG "Statistical Methodologies"	Greg Campbell (CDRH/FDA)
	10. Titration Manuscript(s)	Marc Salit (NIST)
	11. Genome-Wide Association	Huixiao Hong (NCTR/FDA)
	12. One-color *vs*. Two-color: Neuroblastoma (Agilent Platform)	Russ Wolfinger (SAS) Benedikt Brors (DKFZ)
	13. Multiple Myeloma: Gene Expression and Genotyping	Yiming Zhou (UAMS)
	14. Uncertainties in the Multiple-Biomarker Classifier Problem	Weijie Chen (CDRH/FDA)
	15. What Has Been Learned/Improved (TGxDAT Experience)?	Roger Perkins (NCTR/FDA)
	16. Multi-Path Learning Integrates Pathways into Microarray Analysis	Andrej Bugrim (GeneGo)
	… Additional Manuscript Proposals Are Welcome	Volunteers
11:00 am	Coffee Break & Poster View	
Session III-B: Manuscript Preparation (2) Chair: **Leming Shi** (NCTR/FDA)		
11:30 am	Discussion and Presentation of Additional Manuscript Topics	All
12:20 am	Timeline VO: April 28 (Detailed Manuscript Outline) V1: July 14 (Full Manuscript) V2: Aug. 4 (Revised) V3: Aug. 18 (Revised, Ready for Institutional Clearance) V4: Sept. 1 (Revised, Almost Ready for Peer Review) VS: Sept. 8, 2008 (Submission for Peer Review)	Leming Shi (NCTR/FDA)
12:30 pm	Lunch & Poster View	
Session III-C: Genome-Wide Association Working Group (GWAWG) Chair: **Federico Goodsaid** (CDER/FDA)		
1:30 pm	GWA Data Sets	Nick Xiao (SAIC/NCI)
1:50 pm	GWA Data Analysis Plans	Huixiao Hong (NCTR/FDA)
2:20 pm	WTCCC Data Analysis Team	Silvia Vega (Rosetta) Li Zhang (CDER/FDA)
3:00 pm	Genome-Wide CNV Association and Batch Effects in the WTCCC Data Sets	Christophe Lambert (Golden Helix)
3:30 pm	JMP Genomics and GWA Data Analysis	Russ Wolfinger (SAS)
Concurrent Session: MAQC-II Titration Working Group Meeting (Mini Conference Room 1106) *Chair: **Marc Salit** (NIST)*		
1:30 pm ~ 3:30 pm	*Walt Liggett, Jean Lozach, Anne Bergstrom Lucas, Ron Peterson, Rich Shippy, Martin Schumacher, Leming Shi, Jean and Danielle Thierry-Mieg, Russ Wolfinger, ...*	
4:00 pm	Discussion	All
4:50 pm	Summary of the Meeting	Leming Shi (NCTR/FDA)
5:00 pm	Adjourn the Meeting	

Registration

The MAQC meeting is open to everyone and there is no registration fee. However, if you plan to attend the meeting, please contact Leming Shi (Leming.Shi@fda.hhs.gov, +1-870-543-7387) as soon as possible so that a seat will be reserved for you.

Meeting Venue

US Food and Drug Administration
Advisors and Consultants Staff Conference Room
Room 1066
5630 Fishers Lane
Rockville, MD 20857, USA

Poster Presentations

All meeting participants are encouraged to present their data analysis results or other information related to microarrays in posters to enhance the interactions among meeting participants.

Transportation

The local airports are:

Ronald Reagan Washington National Airport (DCA)
Washington Dulles International (IAD)
Baltimore/Washington International Thurgood Marshall (BWI)

The FDA conference room is close to the Metro Twinbrook station of the "red line" (http://www.wmata.com/metrorail/systemmap.cfm).

Hotel

The following hotels are within walking distance to the FDA conference room and to the Metro Twinbrook station of the "red line" (http://www.wmata.com/metrorail/systemmap.cfm):

Hilton Washington DC/Rockville Executive Meeting Center
1750 Rockville Pike
Rockville, MD 20852
301-468-1100
http://www.rockvillehotel.com/

Ramada Inn Rockville
1775 Rockville Pike
Rockville, MD 20852
301-881-2300
http://www.ramadarockville.com/

Crowne Plaza Hotel
Washington DC - Rockville
3 Research Court
Rockville, MD 20850
301-840-0200
This hotel is about 5 miles from the meeting site and offers a free shuttle. The rates start at $180 per night.
http://www.ichotelsgroup.com/h/d/cp/1/en/hotel/rkvrc?&cm_mmc=mdpr-_-yahoosspUS-_-cp-_-rkvrc

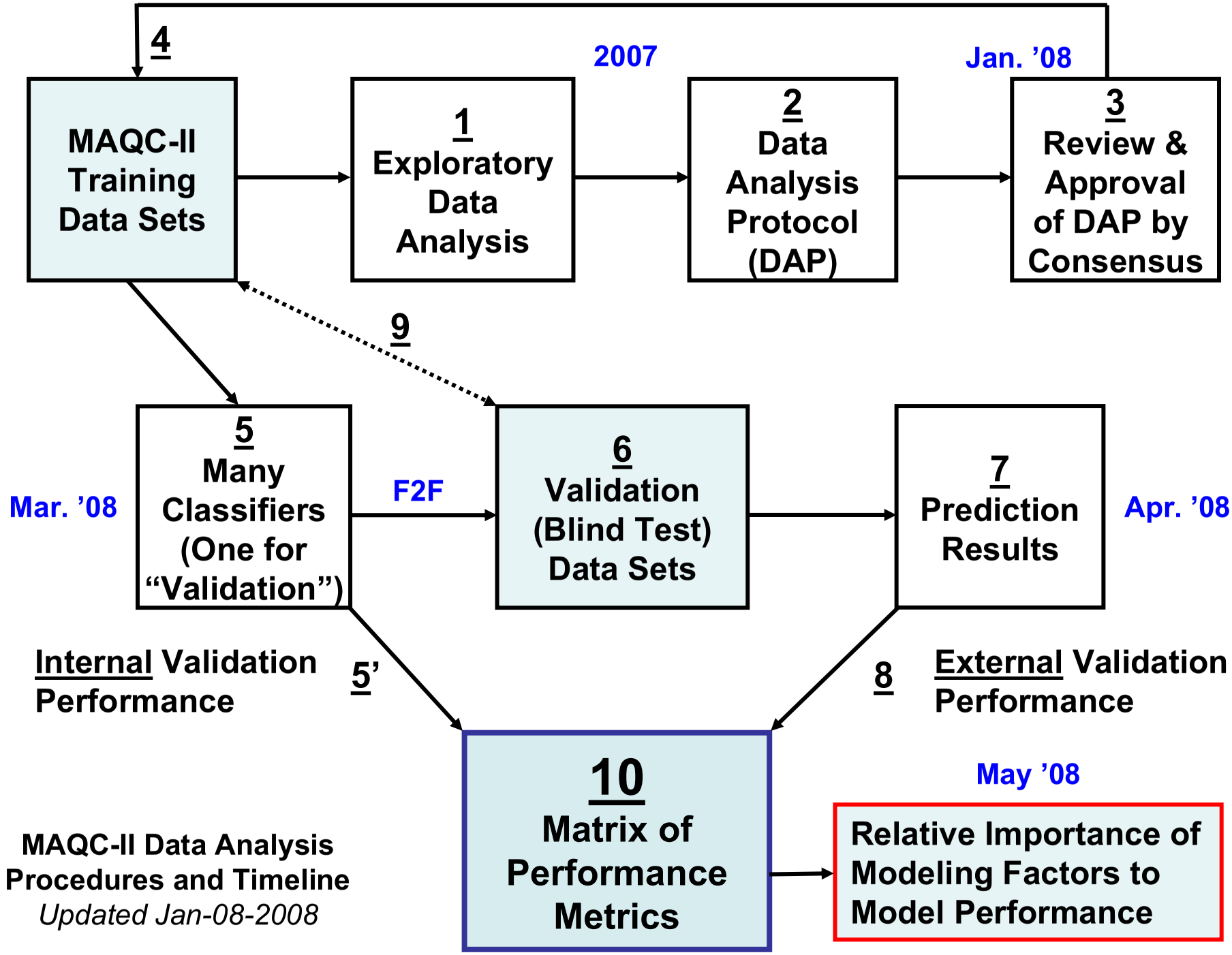
Summary of MAQC-II Data Sets

Leming.Shi@fda.hhs.gov Tel: +1-870-543-7387 Updated: January-17-2007

Table 1 A: MAQC-II Data Analysis Teams Are Required to Predict the 13 Endpoints from Six Data Sets

No.	Date Set Code	Endpoint Code	Endpoint Description	Excel Column Header*	Excel Column*	Number of Samples	Positives	Negatives	P/N Ratio
1	Hamner	A	Lung Tumor	**Class_LT_NLT**	**C**	70	26	44	0.59
2	Iconix	B	Liver Carcinogen	**Class**	**B**	216	73	143	0.51
3	NIEHS	C	Overall Necrosis Score	**Class**	**C**	214	79	135	0.58
4	BR	D	Treatment Response	**pCR**	**O**	130	33	97	0.34
5		E	Estrogen Receptor Status	**erpos**	**H**	130	80	50	1.6
6	MM	F	Overall Survival Milestone Outcome	**OS_MO**	**AB**	**340**	**51**	**289**	**0.18**
7		G	Event-free Survival Milestone Outcome	**EFS_MO**	**AA**	**340**	**84**	**256**	**0.33**
8		H	Clinical Parameter S1	**CPS1**	**S**	340	194	146	1.33
9		I	Clinical Parameter R1	**CPR1**	**T**	340	200	140	1.43
10	NB	J	Overall Survival Milestone Outcome	**OS_MO**	**AM**	**238**	**22**	**216**	**0.10**
11		K	Event-free Survival Milestone Outcome	**EFS_MO**	**AL**	**239**	**49**	**190**	**0.26**
12		L	Newly Established Parameter S	**NEP_S**	**AN**	246	145	101	1.44
13		M	Newly Established Parameter R	**NEP_R**	**AO**	246	145	101	1.44

*See Excel files listed in Table 1B. Endpoints (OS and EFS related) described in red have been updated based on the outcome of a "milestone" survey time; note that these endpoints are unbalanced.


4
2007
Jan. '08
MAQC-II Training Data Sets
1 Exploratory Data Analysis
2 Data Analysis Protocol (DAP)
3 Review & Approval of DAP by Consensus
9
5 Many Classifiers (One for "Validation")
Mar. '08
F2F
6 Validation (Blind Test) Data Sets
7 Prediction Results
Apr. '08
Internal Validation Performance
5'
8
External Validation Performance
10 Matrix of Performance Metrics
May '08
Relative Importance of Modeling Factors to Model Performance
MAQC-II Data Analysis Procedures and Timeline
Updated Jan-08-2008