Summary of Eight (8) Manuscripts to Be Submitted to
The Pharmacogenomics Journal (TPJ)
from the MAQC-II Project

Leming Shi
leming.shi@fda.hhs.gov
Tel: 870-543-7387 (O), 501-200-7311 (M)

November 2, 2009

Microarray technologies are widely used in basic and clinical research including the identification of biomarkers of drug efficacy and safety. In order to fully translate these technologies to the clinic, established clinical and regulatory practices must be followed, which require the development of standards and quality measures.

Initiated by the U.S. Food and Drug Administration (FDA) with about 200 participants from academia, industry, and government, the MicroArray Quality Control (MAQC) project aims to foster the proper application of microarrays in discovery, development and review of FDA regulated products. The first phase (MAQC-I) demonstrated the technical reliability of microarray technology in terms of identifying differentially expressed genes (*Nature Biotechnology* 2006, http://www.nature.com/nbt/focus/maqc/). The second phase (MAQC-II) assessed the capabilities and limitations of various data analysis methods in developing and validating microarray-based classification models for predicting patient outcomes, providing a solid scientific foundation for realizing personalized medicine. These two types of microarray applications (DEGs and predictive models) have been continuously submitted to the FDA by the pharmaceutical and diagnostic industries and require different "best practices" for data analysis and guidance. The issues being addressed by the MAQC-II manuscripts are critically important to the reliable applications of microarrays in the clinic and have not previously been adequately addressed.

Major findings from the MAQC-II have been presented in over 10 manuscripts, a few of which are currently under peer review by *Nature Biotechnology*, also Nature Publishing Group journal. ***Eight manuscripts will be submitted between November 8 and 15, 2009*** to *The Pharmacogenomics Journal (TPJ)*. The MAQC-II project and the FDA are fully committed to presenting MAQC-II results in the highest possible quality publications and would therefore greatly appreciate and incorporate constructive suggestions, comments, and criticisms from the *TPJ* peer reviewers so that the quality of the manuscripts will be further improved. We strongly believe that, as personalized medicine depends on best practices in developing and validating microarray-based predictive models, the findings of the MAQC-II project will be of tremendous interests to the readers of *TPJ* in particular and to the scientific and clinical communities in general.

We have contacted NPG and arrangements have been made for NPG to ***publish a joint NPG reprint (supplemental) issue*** by including all MAQC-II manuscripts that will eventually be published by *NBT* and *TPJ* after peer review. NPG plans to distribute the supplemental issue to

readers of journals such as *Nature Medicine*, *Nature Genetics*, and *Nature Reviews Drug Discovery*. We hope that the eight manuscripts ***will be reviewed by TPJ in an expedited manner*** to ensure that MAQC-II manuscripts submitted to *NBT* and *TPJ* will be published around the same time in order to generate the maximum possible impact.

In the following pages we provide a brief summary of each of the eight manuscripts and hope that each peer reviewer will have a better understanding of the overall scope of the MAQC-II.

1. Enhancing classifier prediction performance by correcting batch effects in microarray gene expression data (Luo J *et al*.)

Batch effects are unwanted, systematic, and non-biological variations between microarray data from batches of samples under various processing conditions such as different sample preparation and hybridization protocols. They are ubiquitous in microarray data and could have disastrous impact on the prediction of patient outcomes from independent cohorts. Using a large number of data sets from the MAQC-II project and various data analysis methods, we demonstrate that batch-effect removal methods can effectively improve the prediction performance of microarray-based predictive models and should be incorporated as a routine practice in analyzing microarray data for prediction purposes. The work also provides a carefully selected compilation of data sets with varying degrees of model predictability for benchmarking future development of batch-effect removal algorithms. We anticipate that the findings and recommendations presented in our manuscript will likely be applicable to enhance classifier performance based on next-generation sequencing data or other types of omic data in which batch effects are also commonly observed.

Keywords: batch effects; MAQC-II; classifier; prediction performance; omics

2. Comparable performance of one-color and two-color gene expression analyses in predicting clinical endpoints of neuroblastoma patients (Oberthuer A *et al*.)

Although the concordance between one-color and two-color microarrays in terms of detecting differentially expressed genes has been demonstrated (e.g., by MAQC-I, NBT 2006), their concordance in performance in terms of predicting patient outcome has not been well demonstrated. To address this important issue that is still under debate, we specifically generated a large data set of almost 500 neuroblastoma samples with both one-color and two-color platforms. We found that predictive models can be generated from one-color and two-color microarray platforms with equivalent prediction performance. Overlap of selected features between both platforms decreases with increasing predictive difficulty of endpoints; however, mechanistically relevant features may be identified by either platform. This data set is expected to be used by the community as an important benchmark set.

Keywords: MAQC-II; microarray; one-color; two-color; neuroblastoma; classification of clinical endpoints

3. Consistency of predictive classifiers and signature genes generated using different microarray platforms (Fan X *et al*.)

Microarray-based classifiers and associated signature genes generated from various platforms are abundantly reported in the literature; however, the utility of the classifiers and signature genes in cross-platform prediction applications remains largely uncertain. As part of the MicroArray Quality Control Phase II (MAQC-II) project, we demonstrate in this study good cross-platform prediction consistency using a large toxicogenomics data set by illustrating that: (1) the signature genes of a classifier generated from one platform can be directly applied to another platform to develop a predictive classifier; (2) a classifier developed using data generated from one platform can accurately predict samples that were profiled using a different platform. The results suggest the potential utility of using published signature genes in cross-platform applications and the possible adoption of the published classifiers for a variety of applications. The study reveals an opportunity for possible translation of biomarkers identified using microarrays to clinically validated assays.

Keywords: microarray; cross-platform; gene signature; classifier; MAQC-II; hepatotoxicity

4. Genomic indicators in the blood predict a drug-induced liver injury response (Huang J *et al*.)

Genomic biomarkers for the detection of drug-induced liver injury (DILI) from blood are urgently needed for monitoring drug safety. We used a unique data set from the Food and Drug Administration led MicroArray Quality Control Phase II (MAQC-II) project consisting of gene expression data from the two tissues (blood and liver) to test cross-tissue predictability of genomic indicators to a form of chemically-induced liver injury. We then use the genomic indicators from the blood as biomarkers for prediction of acetaminophen-induced liver injury and show that the cross tissue predictability of a response to the pharmaceutical agent (accuracy as high as 92.1%) is better than, or at least comparable to, that of non-therapeutic compounds. We provide a database of gene expression for the highly informative predictors which brings biological context to the possible mechanisms involved in DILI. Pathway-based predictors were associated with inflammation, angiogenesis, Toll-like receptor signaling, apoptosis and mitochondrial damage. The results demonstrate for the first time and support the hypothesis that genomic indicators in the blood can serve as potential diagnostic biomarkers of DILI.

Keywords: prediction; acetaminophen; blood; cross tissue; liver injury; microarray gene expression

5. A novel feature selection for microarray data analysis and its applications to the MAQC-II data and two additional breast cancer data sets (Cheng J *et al*.)

We propose a novel method for selecting predictive features from high dimensional genomic data. The method utilizes two-way filtering to find the most informative features regardless of

feature distribution characteristics. We demonstrate the utility of the method on six MicroArray Quality Control Phase II (MAQC-II) datasets. The results indicate that our method yields models with fewer features and can achieve comparable or superior classification performance compared to models generated from other feature selection methods. This suggests that our method is very effective in identifying a small number of highly informative features. We also tested the potential clinical value of the method by applying it to challenging clinical classification problems including treatment response and prognostic prediction in clinically homogeneous subsets of breast cancers. We identified candidate markers of response to chemotherapy in triple receptor-negative patients and prognostic markers for patients stratified by clinical variables such as estrogen receptor status.

Keywords: biomarker discovery; gene expression data analysis; breast cancer; tumor treatment response; tumor prognosis; MAQC-II

6. K-nearest neighbors (KNN) models for microarray gene-expression analysis and reproducible clinical outcome prediction (Parry RM *et al*.)

A number of factors contribute to the reproducibility and reliability of prediction performance in genomic data analysis. This paper focuses on predictive models using the K-nearest neighbor (KNN) methodology because it has broad interest within medical and science communities, and has shown large performance variations among multiple KNN models in the MAQC-II project. MAQC-II provides a unique set-up of 10 clinical endpoints from clinical breast, neuroplastoma, multiple myeloma cancers consisting of ~700 samples, and positive and negative control data sets consisting of 496 samples. Because KNN structure is simple, we systematically evaluate 463,320 models, and identify the factors that contribute to the variations. We propose a KNN-based sensible data analysis protocol (DAP) and identify candidate model parameters achieving good cross-validation performance. We show that the sensible KNN-based DAP achieves reliable prediction of clinical outcomes (i.e. external validation) by using a newly generated dataset of 478 neuroblastoma patients (147 words).

Keywords: cancer prediction; KNN predictor; parameter selection; reproducibility; sensible data analysis protocol; MAQC-II

7. Functional analysis reveals the biological underpinnings of predictive genomic signatures (Shi W *et al*.)

Gene expression signatures of toxicity and clinical response benefit both safety assessment and clinical practice. Difficulties in understanding the association of the signatures to the predicted endpoints have, however, limited their application. The MAQC-II project generated 262 signatures for ten clinical and three toxicological endpoints from six gene expression data sets. A comprehensive functional analysis of these signatures and their non-redundant unions was conducted using ontology enrichment, biological network building and interactome connectivity analyses. Different signatures for a given endpoint were more similar at the level of biological entities and transcriptional control than at the gene level. Signatures tended to be enriched in

function and pathway in an endpoint and model-specific manner, and showed a topological bias for incoming interactions. Importantly, the level of biological similarity between different signatures correlated positively with the accuracy of the signature predictions. These findings will impact the design, understanding, and application of predictive genomic signatures and support their broader application in predictive medicine.

Keywords: Genomic signatures; enrichment analysis; network reconstruction; biological pathways; interactome; MAQC-II

8. Maximizing biomarker discovery by minimizing gene signatures (Chang C *et al*.)

Many approaches have been developed for identifying gene expression biomarkers to predict clinical outcomes, but there is a lack of systematic evaluation of the advantages and limitations of different approaches. Here, we have used multiple level similarity analyses for breast cancer endpoints by comparing classifier models generated from dozens of data analysis teams of the MicroArray Quality Control Phase-II (MAQC-II) Consortium to explore the consistency and diversity of different predictive biomarkers. We propose an innovative method termed Minimum Feature Size (MFS), which enables gene expression biomarker selection with reduced redundancy and also improves prediction performance based on biological function. Our method is demonstrated on two well-known breast cancer data sets and the results suggest that classifiers with smaller feature sizes exhibit improved performance during internal and external validation. The implementation of our method on the biomarker and gene signature selection can potentially lead to a higher cost-effectiveness in the clinical application.

Keywords: MAQC-II; gene expression; biomarkers; similarity analyses; breast cancer