The MAQC Project: Calibrated RNA Samples, Reference Datasets, and QC Metrics/Thresholds for Microarray Quality Control

Teleconference Date: April 06, 2006 (9 am PDT/11 am CDT/12 pm EDT/16:00 GMT)
Summary Date: April 06, 2006
Author: Leming Shi, +1-870-543-7387, Leming.Shi@fda.hhs.gov
MAQC Website: http://edkb.fda.gov/MAQC/ (Find all meeting summaries, *etc*.)

1. Summary of Apr-03 TC on MS-3 (main paper):

Representatives from official data analysis sites and platform providers, and MS-3 drafting team members attended a conference call on the main paper on Monday, April 3. We covered many topics and agreed on many issues during the 2.5 hrs' discussions:

(1) We agreed that it was reasonable to remove obvious outlier arrays from analysis when quality control criteria to identify such outliers are provided and stated in the manuscript. Each array manufacturer was asked to look at its data carefully and to report possible outlier arrays to Leming Shi for by 10 pm CST, Tuesday, April 4.
(2) For the calculation of cross-site intensity CV (Fig. 3A bottom plot, V1 draft of Mar-19-06), global scaling across sites was requested by ILM and was agreed upon by the group. This additional global scaling will only be applied to Fig. 3A (bottom) and will be mentioned in the manuscript.
(3) A few figures will be removed: Fig. 3B, Fig. 4, and Fig. 6.
(4) Ratio scatter plots will be added for each platform relative to TAQ (p.12 bottom). Catalin Barbacioru will also be creating some figures for the section on "Comparison to TAQ" (p.12).
(5) We'll decide a good way to present cross-platform comparability (p.12).
(6) At the suggestion of John Corson, power analysis (p.13) will be redone. Charlie Johnson and John Corson will be working on this.

2. Resolving potential overlaps between MS-3 and other manuscripts (*e.g.*, MS-7):

Rick and Wendell share some sample scatterplots with the group and we like them. However, concerns were raised on the potential redundancy between MS-3 and MS-7 regarding the scatterplot comparison of microarray data with TAQ, QGN, and GEX. One suggestion was to summarize such comparison results in a compact table format while leaving the scatterplots in MS-7. However, several array manufacturers expressed preference of putting these scatterplots in MS-3. We'll create the figures first (based on the final data to be used in analysess, see point 4 below) and then decide where they may best fit.

3. Outlier arrays to be excluded from analysis:

AG1 identified four outlier arrays (AG1_1_A1, AG1_2_A3, AG1_2_D2, AG1_3_B3) and ABI identified two outlier arrays (ABI_2_A5 and ABI_3_D2). These six outlier arrays should be eliminated from all analyses and the reasons (*i.e.*, based on manufacturers' internal QC criteria) for their exclusion will be clearly stated in the

manuscript. Detailed information is available from Jim Collins (AG1) and Yongming Sun (ABI) or Leming Shi upon request.

4. Data sets for final analyses – different N's will be reported (No universal subset of "Present" genes):

We agreed that for most analyses (figures/tables) the entire mapped set of 12,091 genes should be used. However, when CV or FC (fold change) is calculated and compared, "Present" rules (*i.e*., "Present" in the majority of replicates for both samples/sites/platforms) should be applied. We decided not to create a universal "Present" subset of genes for analyses. Instead, different N's in each figure/table for each comparison will be reported based on the majority rule of "Present". The Y-axis of power analysis will show the number of genes instead of %.

5. Avoiding potential perception of biased analyses:

Leming Shi emphasized the importance of avoiding potential negative perceptions due to biased analyses. We have generated wonderful data within the MAQC project and we do not had to analyze our data in a way that might leave the scientific community the perception that we tried to "massage" our data to let them LOOK good - we should not and do not need to do that. Equally importantly, we should also avoid obvious pitfalls in microarray data analysis (*e.g*., focusing on noise).

6. "April 10" deadline extended to "April 20":

We agreed to extend the "April 10" deadline to "April 20" for the submission of draft version 2 to be distributed to many organizations for internal review. The extension was agreed upon based on the understanding that it will absolutely not delay the "May 10" deadline for draft version 3.

7. Mapping paper:

At *Nature Biotechnology*'s suggestion, there will be only one mapping paper - MS-5A (led by Jean). Jean welcomes contributions from members of the MAQC group.

8. Sanity check:

Wendell, Rich, and Damir will cross-check their counts of genes to be used for the analyses.

9. Others:

Laura Reid will be updating MS-3 draft.
We'll have another MS-3 conference call on Friday, April 14.

Next MAQC Teleconference:
Thursday, April 20, 2006 (9 am PDT / 11 am CDT / 12 pm EDT / 16:00 GMT)
USA Toll Free Number: **888-566-5020**; International caller: +1-210-795-9594
PASSCODE: **79451**