

# The MAQC (Microarray Quality Control) Project: Calibrated RNA Samples, Reference Datasets, and QC Metrics and Thresholds

Leming Shi<sup>1</sup>, Felix W. Frueh<sup>2</sup>, Uwe Scherf<sup>3</sup>, Raj K. Puri<sup>4</sup>, Scott A. Jackson<sup>5</sup>, Heather C. Harbottle<sup>6</sup>, Janet A. Warrington<sup>7</sup>, Jim Collins<sup>8</sup>, David Dorris<sup>9</sup>, Lu Zhang<sup>10</sup>, Laurence Lamarca<sup>11</sup>, Yaping Zong<sup>12</sup>, Timothy J. Sendera<sup>13</sup>, Shawn C. Baker<sup>14</sup>, Gavin M. Fischer<sup>15</sup>, Bud Bromley<sup>16</sup>, David J. Dix<sup>17</sup>, Marc Salit<sup>18</sup>, Zoltan Szallasi<sup>19</sup>, Ernest S. Kawasaki<sup>20</sup>, Charles Wang<sup>21</sup>, Roderick V. Jensen<sup>22</sup>, Jing Han<sup>4</sup>, Federico M. Goodsaid<sup>2</sup>, Gene A. Pennello<sup>3</sup>, Thomas A. Cebula<sup>5</sup>, Joseph E. LeClerc<sup>5</sup>, Weida Tong<sup>1</sup>, James C. Fuscoe<sup>1</sup>, Tucker A. Patterson<sup>1</sup>, Tao Han<sup>1</sup>, Lei Guo<sup>1</sup>, Hong Fang<sup>1</sup>, James J. Chen<sup>1</sup>, Yvonne P. Dragan<sup>1</sup>, William Slikker, Jr.<sup>1</sup>, Daniel A. Casciano<sup>1</sup>. <sup>1</sup>NCTR, FDA, Jefferson, AR; <sup>2</sup>CDER, FDA, Rockville, MD; <sup>3</sup>CDRH, FDA, Rockville, MD; <sup>4</sup>CBER, FDA, Bethesda, MD; <sup>5</sup>CFSAN, FDA, Laurel, MD; <sup>6</sup>CVM, FDA, Laurel, MD; <sup>7</sup>Affymetrix, Santa Clara, CA; <sup>8</sup>Agilent, Palo Alto, CA; <sup>9</sup>Ambion, Austin, TX; <sup>10</sup>Applied Biosystems, Foster City, CA; <sup>11</sup>BD Clontech, Palo Alto, CA; <sup>12</sup>Full Moon BioSystems, Sunnyvale, CA; <sup>13</sup>GE Healthcare, Chandler, AZ; <sup>14</sup>Illumina, San Diego, CA; <sup>15</sup>Stratagene, La Jolla, CA; <sup>16</sup>ViaLogy, Altadena, CA; <sup>17</sup>NHEERL, EPA, Research Triangle Park, NC; <sup>18</sup>NIST, Gaithersburg, MD; <sup>19</sup>Harvard, Boston, MA; <sup>20</sup>NCI, Bethesda, MD; <sup>21</sup>UCLA, Los Angeles, CA; <sup>22</sup>UMass, Boston, MA. [Leming.Shi@fda.hhs.gov](mailto:Leming.Shi@fda.hhs.gov)

## ABSTRACT

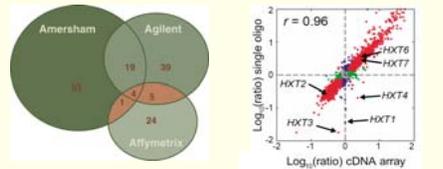
FDA's Critical Path Initiative identifies pharmacogenomics and toxicogenomics as key opportunities in advancing medical product development and personalized medicine, and the "Guidance for Industry: Pharmacogenomic Data Submissions" has been released. Microarrays represent a core technology in pharmacogenomics and toxicogenomics; however, before this technology can successfully and reliably be used in clinical practice and regulatory decision-making, standards and quality measures need to be developed.

The Microarray Quality Control (MAQC) project currently involves six FDA Centers, major providers of microarray platforms and RNA samples, EPA, NIST, academic laboratories, and other stakeholders. The MAQC project aims to establish QC metrics and thresholds for objectively assessing the performance achievable by various microarray platforms and evaluating the advantages and disadvantages of various data analysis methods. Two RNA samples will be selected for three species, human, rat, and mouse, and differential gene expression levels between the two samples will be calibrated with both microarrays and QRT-PCR. The resulting microarray datasets will be used for assessing the precision and cross-platform/laboratory comparability of microarrays, and the QRT-PCR datasets will enable evaluation of the nature and magnitude of any systematic biases that may exist between microarrays and QRT-PCR. The availability of the calibrated RNA samples combined with the resulting microarray and QRT-PCR datasets, which will be made readily accessible to the microarray community, will allow individual laboratories to more easily identify and correct procedural failures. The MAQC project will help improve the microarray technology and foster its proper applications in discovery, development and review of FDA regulated products.

Everyone is invited to participate in the MAQC project.

<http://edkb.fda.gov/maqc/>

Views expressed in this presentation are those of the presenters and not necessarily those of the U.S. FDA.



E. Marshall, *Science* 306, 630 (2004);  
P.K. Tan et al., *Nucleic Acids Res* 31,  
5676 (2003).

Hughes TR et al., *Nature Biotechnology*,  
19(4), 342 (2001)

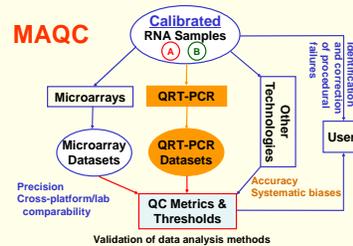
**Figure 1. Conflicting results have been reported in the literature regarding the cross-platform comparability, and, hence, the reliability of microarray technology.**

## The Microarray Community Is Facing Two Major Challenges:

- How to ensure the experimental proficiency of individual laboratories in comparison to the achievable performance of the microarray technology
- How to objectively assess the merits of various data analysis methods

Because there is a lack of

- Calibrated RNA samples
- Reliable benchmark datasets



**Figure 2. An Overview of the MAQC Project**

## Selection of RNA Samples (A) (B)

Two RNA samples for each species (Human, Mouse, and Rat)

Starting with one species (Human).

Criteria for RNA sample selection:

1. Available in large quantity
2. Reproducibility in production
3. High quality
4. Accessibility (commercial sources)
5. Wide gene presence
6. Large fold changes for a number of genes

Options for RNA sample selection:

1. Two universal reference RNAs
2. Two tissue-specific RNAs
3. Two cell lines
4. Combination

## The MAQC Pilot Study: Selecting Two RNA Samples for the MAQC Main Study

Four Candidate RNA Samples:

- A. Ambion Brain RNA
- B. Ambion Liver RNA
- C. Clontech UHRR (Universal Human Reference RNA)
- D. Stratagene UHRR (Universal Human Reference RNA)

Four Platforms:

1. Affymetrix
2. Agilent
3. GE Healthcare
4. Illumina

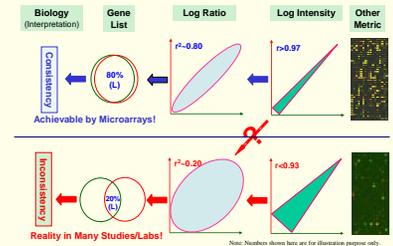
Five (5) replicates per sample per test site for one-channel platforms, resulting in 20 hybridizations per site per platform.

Six Test Sites (7 datasets):

1. Affymetrix (Affymetrix)
2. Agilent (Agilent)
3. Ambion (Affymetrix and GEHC)
4. Illumina (Illumina)
5. NCTR (Agilent)
6. UMass Boston (GEHC)

160 hybridizations

For Agilent platform, 6 sample-pairs were performed in 5 replicates, resulting in 30 hybridizations per test site (Agilent and NCTR).



**Figure 3. Establishing microarray QC metrics and thresholds**

## CONCLUSIONS

**Advantages of Calibrated RNA Samples:**

- Consistent profiling of an identical pair of samples
- Data sharing and comparison made possible
- Establishment of Standard Operating Procedures (SOPs)
- Quality control/assurance (preventing procedural failures)

**Benefits of Reference Datasets:**

- Achievable performance on each platform
- Precision, reproducibility, accuracy, and robustness
- Cross-laboratory and cross-platform concordance
- QC metrics and thresholds
- Evaluation of the merits of data analysis methods

**Guidelines for Regulatory Review of PGx Data Submissions**

## ArrayTrack:

MAQC Data Centralization/Distribution

MAQC Pilot Study datasets (160 arrays) have been submitted to NCTR. Data were centralized within ArrayTrack and distributed to 11 sites selected for performing independent data analysis of the 7 datasets:

1. Affymetrix
2. Agilent
3. Ambion
4. Applied Biosystems
5. Clontech
6. GE Healthcare
7. Illumina
8. NCTR
9. NIST
10. Stratagene
11. UMass Boston

<http://edkb.fda.gov/webstart/arraytrack/>