



The MicroArray Quality Control (MAQC) Project

MAQC Main Study Guidance

Appendix 5: MAQC Guidance to Data Analysis Version 3

Leming Shi

National Center for Toxicological Research (NCTR)

U.S. Food and Drug Administration (FDA)

3900 NCTR Road, Building 5C / Room 109L

Jefferson, Arkansas 72079, U.S.A.

+1-870-543-7387

Leming.Shi@fda.hhs.gov

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

<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/>

November 14, 2005



Table of Contents

1. Summary 4

2. Manufacturer’s Preferred Procedures for Data Preprocessing and Normalization 4

 2.1 Affymetrix (AFX) 5

 2.2 Agilent (AGL) 6

 2.3 Applied Biosystems (ABI) 7

 2.4 Eppendorf (EPP) 8

 2.5 GE Healthcare (GEH) 9

 2.6 Illumina (ILM) 9

 2.7 NCI_Operon (NCI) 9

 2.8 TaqMan (TAQ) 10

 2.9 QuantiGene (QGN) 10

 2.10 StaRT-PCR (GEX) 10

 2.11 Summary of Platform-specific Data Treatment 11

 2.12 Unique Probe ID Column(s) for Each Platform 12

 2.13 Manufacturer’s Contact Person 12

3. Data Management with ArrayTrack 12

 3.1 Annotation Information Provided by Manufacturers 13

 3.2 Data Columns Imported to ArrayTrack 13

 3.3 Data Distribution for Independent Analysis 14

4. Microarray Quality Control Metrics and Thresholds 15

 4.1 QC Metrics Based on a Single Array 16

 Single-array QC Metrics (Before Hybridization) 16

 Single-array QC Metrics (After Hybridization) 16

 4.2 QC Metrics Based on Replicate Arrays 16

 4.3 QC Metrics Based on a Study 17

 4.4 QC Metrics on Accuracy 18

 4.5 QC Metrics for Cross-site Comparison 19

 4.6 QC Metrics for Cross-platform Comparison 19

 4.7 QC Thresholds 20

5. Criteria for Defining Subsets of Genes for Cross-platform Comparison 20

 5.1 Matching of RefSeq IDs Based on Manufacturers’ Annotation 20

 5.2 Mapping of Probe Sequences to RefSeq Transcripts 21

 5.3 Handling of Probe-Transcript (Target) Relationships 21

 1. One Probe – One Transcript (Single) 21

 2. Multiple Probes – One Transcript (Duplicate) 21

 3. One Probe – Multiple Transcripts (*Multiple*) 22

 4. Multiple Probes – Multiple Transcripts (*Multiple*) 22

 5. Contact Person for the Probe–Target Mapping Efforts 23

 5.4 Handling of Flags and Low-intensity Transcripts 23

6. Data Analysis Considerations 23

 6.1 Outlier Arrays 23

 6.2 Detection of Expressed Genes - Present Calls 23

 6.3 Detection of Differentially Expressed Genes - Gene Selection Methods 23

 6.4 Normalization Methods 24



6.5 Exploratory Analysis Methods 24

6.6 Correlation between Array Data Quality and the Quality of RNA, Target, etc..... 25

7. Comparison of Microarray with TaqMan, QuantiGene, and StaRT-PCR..... 25

8. Analysis of Titration Datasets..... 25

9. Performance of Spike-in Controls 27

10. Comparison of Scanner Performance 27

11. Identification of Genes Showing the Largest Discordance..... 27

12. Presentation of Analysis Results at the December 1-2, 2005 Meeting..... 27

13. Planning for Publications 27

14. References..... 28

15. Web Sites 29

Major changes since draft version 2 of September 21, 2005:

- CMB (Combimatrix) decided to withdraw from the MAQC study; thus descriptions about CMB have been removed.
- Updated the platform providers’ preferred data preprocessing and normalization procedures.
- Some degree of consensus has been reached in handling probe-target mapping.
- Yuling Luo (Genospectra) provided a description on how to analyze titration data to assess accuracy.
- Proposed a more detailed agenda for the MAQC December meeting in Palo Alto, CA.
- Proposed a more detailed plan for MAQC publications.
- This version of the draft (v3) will only be distributed among the test sites and data analysis sites that have agreed to the conditions of confidentiality in accessing and analyzing the MAQC datasets.

1. Summary

Data analysis remains one of the biggest challenges in microarray-based studies. There is a lack of guidance for microarray data analysis since this is a fast evolving field. Nevertheless, it is imperative that some general guidelines be set for the analysis of the MAQC study datasets in order to achieve the intended goals of the study.

The MAQC Guidance to Data Analysis is the 5th Appendix of the MAQC Main Study Guidance (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/>) and describes:

- Manufacturer's preferred procedures for data preprocessing and normalization;
- Quality control metrics and thresholds;
- Sequence-based cross-platform mapping;
- Criteria for defining subsets of genes for cross-platform comparison;
- Comparison of microarrays with alternative technologies;
- Identification of the most discordant genes;
- Presentations at the December 1-2, 2005 meeting in Palo Alto, CA;
- Planning for publications.

Everything described in this document is intended for stimulating discussions within the MAQC group and not intended for excluding participants from analyzing the data in their favorable ways. Please send your comments, suggestions, and corrections to Leming.Shi@fda.hhs.gov.

2. Manufacturer's Preferred Procedures for Data Preprocessing and Normalization

FDA/NCTR receives microarray datasets from 24 test sites on seven platforms (AFX, AGL, EPP, ABI, GEH, ILM, and NCI) and loads them into ArrayTrack, which is being used by the FDA for the management, analysis, and interpretation of Voluntary Pharmacogenomic Data Submissions (VGDS) from the pharmaceutical industry (<http://www.fda.gov/nctr/science/centers/toxicoinformatics/arraytrack/>).

Each microarray manufacturer participating in the MAQC main study was given an opportunity to provide its preferred procedures for preprocessing and normalizing data generated on its platform. Whenever possible, the datasets to be distributed to test sites and analysis sites are being processed according to manufacturer's preferred procedures. Since the original intensity data will also be distributed, each platform provider and analysis site can verify whether the output from ArrayTrack is correct or not, and can always refer to the original data when needed.

For clarity of description, the four RNA samples used in the MAQC main study are listed in Table 1. Each sample is processed in five replicates at each of the three test sites per

microarray platform except for Agilent (only samples A and B are processed) and Affymetrix (six test sites are involved).

Table 1. Four RNA Samples

RNA	Description	Product number
A	Stratagene Universal Human Reference RNA (SUHRR)	740000
B	Ambion Human Brain Reference RNA (AHBRR)	6050; 6051; 6052
C	75% A and 25% B	N/A
D	25% A and 75% B	N/A

Information about sample A can be found at <http://www.stratagene.com/manuals/740000.pdf>;

Information about sample B can be found at <http://www.ambion.com/catalog/CatNum.php?6050>.

2.1 Affymetrix (AFX)

CEL file (probe level data) is the standard output format of Affymetrix's GeneChip[®] microarray platform and will be distributed along with gene (probe set) level summary data. Affymetrix prefers using PLIER (Probe Logarithmic Intensity Error estimation) to derive probe set level data from CEL files:

1. Run PLIER with the default settings. Make sure that the probe level "Quantile normalization" and "PM-MM" options are set.
2. After PLIER, a constant of 16 is added to the signal value of each probe set for stabilizing variance for probe sets with low intensity.
3. Log₂ transform the data.

Like dCHIP, RMA, and their variants, PLIER is a multi-array normalization method. That is, an array is not considered independent of other arrays; the signal value for a probe set in an array depends not only on the array itself but also on other arrays being used as a group to derive the error model and generate the PLIER probe set level data. For the MAQC main study, the 20 arrays (five replicates for each of the four RNA samples) from each test site are considered as an independent group for generating PLIER output.

PLIER is available in various software packages, *e.g.*, ArrayAssist Lite (http://www.affymetrix.com/products/software/specific/arrayassist_lite.affx) and Bioconductor (<http://www.bioconductor.org/>). ArrayTrack stores CEL files and generates PLIER output by calling the `justplier` function in Bioconductor with the settings preferred by Affymetrix. The PLIER outputs from ArrayTrack, PLIER SDK, and Bioconductor (R package) have been compared and found consistent. However, the FDA/NCTR noticed significant differences between ArrayAssist Lite's PLIER output and that from PLIER SDK or Bioconductor or ArrayTrack. Such discrepancies have been reported to Affymetrix and Stratagene (the developer of ArrayAssist Lite). Consequently, Affymetrix has decided to use Bioconductor to generate PLIER output for the MAQC study. A constant of 16 has been added to the data distribution before log₂ transformation. Flags (P/M/A) are based on the MAS5 algorithm.

For more information about PLIER, visit http://www.affymetrix.com/corporate/events/webtalk_archive_2005.affx ("Coming to Grips

with the PLIER Algorithm”) or “Some M-Estimators for Expression Analysis” by Earl Hubbell (http://www.affymetrix.com/corporate/events/seminar/microarray_workshop.affx).

2.2 Agilent (AGL)

Data are provided in the form of a tab-delimited text (.txt) file that contains three tables of input parameters and output results. These tables are FEPARAMS, STATS, and FEATURES. The FEPARAMS table contains input parameters and options used to run Feature Extraction software. The STATS table gives results derived from statistical calculations that apply to all features on the microarray. The last table, FEATURES, displays results for each feature in 95 output columns, such as gene name, log ratio, processed signal, mean signal, or dye-normalized signal. These columns are described in detail in the user manual.

For the Agilent platform (two-color), only two RNA samples are processed (A – Strategene UHRR and B – Ambion Brian reference RNA). Five replicate hybridizations are processed for each of the four RNA pairs (Table 2):

Table 2. Agilent’s Four RNA Pairs

RNA Pair	RNA Pair Description	Comment
A	SUHRR-Cy3 vs. SUHRR-Cy5	Self-self
B	AHBRR-Cy3 vs. AHBRR-Cy5	
C	SUHRR-Cy3 vs. AHBRR-Cy5	Dye-swap
D	AHBRR-Cy3 vs. SUHRR-Cy5	

Hybridizations from RNA pairs A and B will be used to assess dye biases and false positive rate. RNA pairs C and D will be used to calculate log ratio between SUHRR and AHBRR, *i.e.*, $\log_2(\text{AHBRR}/\text{SUHRR})$.

1. Calculate Detectable Signal Thresholds for red and green channels per microarray:

$$r_{\text{DetectableSignalThreshold}} = (\text{AddErrorEstimateRed}/g_{\text{LinearDyeNormFactor}}) \times 2$$

$$g_{\text{DetectableSignalThreshold}} = (\text{AddErrorEstimateGreen}/r_{\text{LinearDyeNormFactor}}) \times 2$$

- These represent 2 times the error in the feature signal intensities that are close to background for each channel.

- Only background subtracted signal intensities per channel above these thresholds should be considered as detected.

[Fields used from .txt file:

AddErrorEstimateRed additive error component, red channel (STATS Column 124)

AddErrorEstimateGreen additive error component, green channel (STATS Column 123)

rLinearDyeNormFactor = global dye normalization factor, red channel (STATS Column 61)

gLinearDyeNormFactor = global dye normalization factor, green channel (STATS Column 57)]

2. Filter out data points that do not meet the following criteria:

- a. $rBGSubSignal > rDetectableSignalThreshold$ **OR** $gBGSubSignal > gDetectableSignalThreshold$
- b. $ControlType = 0$
- c. $rIsFeatNonUnifOL$ **AND** $gIsFeatNonUnifOL = 0$
- d. $rIsSaturated$ **AND** $gIsSaturated = 0$

A combined flag column (Pass/Fail) was generated for each array by Jim Collins. Only those spots labeled with “Pass” should be used in data analysis.

[Fields used from .txt file:

$rBGSubSignal$ = background subtracted red intensity (FEATURES Column 63)

$gBGSubSignal$ = background subtracted green intensity (FEATURES Column 62)

$ControlType$ = biological probes are indicated as 0; control probes are indicated as 1 or -1 (FEATURES Column 9)

$rIsFeatNonUnifOL$ = feature pixel uniformity outlier, red channel (FEATURES Column 54)

$gIsFeatNonUnifOL$ = feature pixel uniformity outlier, green channel (FEATURES Column 53)

$rIsSaturated$ = saturated feature, red channel (FEATURES Column 50)

$gIsSaturated$ = saturated feature, green channel (FEATURES Column 49)]

3. Calculate ratios as $rProcessedSignal/gProcessedSignal$.

[Fields used from .txt file:

$rProcessedSignal$ = normalized red intensity (FEATURES Column 24)

$gProcessedSignal$ = normalized green intensity (FEATURES Column 23)]

4. Log₂ transform data.

NOTE: LogRatio values output in .txt file (FEATURES Column 16) are Log base 10.

5. LogRatio base 2 values for the dye-swap pair hybridizations should be averaged, and the resulting averaged value is considered as the LogRatio for one of the five replicate pairs.

2.3 Applied Biosystems (ABI)

1. A .txt (or .csv) file contains all arrays with the following columns:

- 1) ProbeID or Probe_Name

The non-numeric ProbeIDs are the control probes and should be removed for the data analysis.

- 2) GeneID or Gene_Name

The corresponding GeneIDs for the control probes are either ‘NULL-GENE’ or blank with no values.

- 3) One Signal column for each array

This is the chemiluminescent intensity with background already subtracted.

- 4) One “S/N” column for each array

This is signal-to-noise ratio value and is used to quantify the likelihood that the chemiluminescent signals is present (above noise). A S/N value above or equal to 3 indicates that the probe is detectable, while a S/N value less than 3 indicates



that the probe is non-detectable. *The S/N column is used to generate the “Flag” information within ArrayTrack in the distributed normalized data.*

- 5) One “Flags” column for each array
A numeric code for the quality of the probe image quantification. A value above 5,000 indicates that the image spot has potential quality issues and cautions should be taken when interpreting analysis results for such probes. *Only a very limited number of spots are above the 5,000; therefore, the “Flags” column is ignored in ArrayTrack data input.*
 - 6) Possibly additional columns for each array
2. For the purpose of MAQC data analysis, only the columns 1-5 listed above will be needed.
 3. Quantile normalization across the 20 arrays per test site is recommended. The normalization should be performed after removing control probes.
 4. Log₂ transform the data after quantile normalization.

2.4 Eppendorf (EPP)

For a more detail description on Eppendorf data format and data analysis procedure, please refer to the “Description of the Data Format from Eppendorf DualChip Microarrays” (provided with EPP data distribution). There are 294 genes printed on each EPP array. Briefly, each of the 20 arrays (replicates) has been “normalized” against a reference sample according to the following pairing (Table 3):

Table 3. Eppendorf’s Sample Pairs

No.	Replicate	Sample Array	Reference Array
1	A1	A1	A1
2	A2	A2	A1
3	A3	A3	A1
4	A4	A4	A1
5	A5	A5	A1
6	B1	B1	A1
7	B2	B2	A2
8	B3	B3	A3
9	B4	B4	A4
10	B5	B5	A5
11	C1	C1	A1
12	C2	C2	A2
13	C3	C3	A3
14	C4	C4	A4
15	C5	C5	A5
16	D1	D1	A1
17	D2	D2	A2
18	D3	D3	A3
19	D4	D4	A4
20	D5	D5	A5

2.5 GE Healthcare (GEH)

Each array is saved in a separate .txt file with many columns.

1. Use the **Normalized_intensity**, which is the result of linearly scaling **Raw_intensity** (*i.e.*, **Spot_mean – Bkgd_median**) to a median of 1 per array.
2. Exclude negative values. It is best to only use spots flagged as “G” so that negative values are automatically excluded.
3. Log₂ transform the normalized intensity for each probe.

Definition of GEH's original flag information (CODELINKFLAG):

G = “Good”

L = “Absent” (below noise since the signal mean is less than background mean plus 1.5 standard deviations of local background)

S = “Saturated”

M = “Manufacturing Defect”

I = “Irregular Shape”

C = “Contaminated Spot”

Only use spots which are flagged as “G” or “L”.

2.6 Illumina (ILM)

1. Data for the 20 arrays per test site are provided in three .csv files: (A) raw data (un-normalized); (B) quantile normalized data; and (C) data after quantile normalization and offsetting minimum signal value to 1 across the arrays from the same test site. Four columns of data are associated with each array: **AVG_Signal**, **BEAD_STDEV**, **AVG_NBEADS**, and **Detection** (≥ 0.99 as “Present”, otherwise “Absent”).
2. Use normalized signal (contained in **AVG_Signal** output column) which is a product of quantile normalization with all arrays from the same test site normalized to the averaged profile of the five samples from group “B”. Since the normalization is applied after background estimated from a set of about 1600 negative controls is subtracted, half of the unexpressed targets is expected to have negative signal.
3. For assessing fold changes use only targets which are declared to be detected (value in **Detection** output column ≥ 0.99).
4. In order to compare signals to log transformed signals from other platforms, add a constant offset to the data. The value of offset is $(1 - \text{MINVAL})$ where **MINVAL** is a minimum target signal observed across all samples.

2.7 NCI_Operon (NCI)

Each of the four RNA samples is processed in five replicates using a common reference design in which Stratagene’s UHRR is labeled with Cy5 as the reference and the four samples are labeled with Cy3. Arrays are scanned using Axon’s GenePix 4000B scanners. The following columns of the GenePix GPR files are imported to ArrayTrack:

F635 Median (Cy5 median intensity)

B635 Median (Cy5 median background)
 F532 Median (Cy3 median intensity)
 B532 Median (Cy3 median background)
 Flags (Flags, “0” indicates good spots)

1. For each channel, the background is subtracted from the spot intensity. If the background-subtracted intensity is not positive then the spot is omitted from further analysis and from the results. Negatively flagged spots (according to GenePix) are ignored throughout the entire process, as if they were never part of the input data.
2. Background-subtracted intensity values are normalized using Lowess.
3. log₂ transform intensity data.
4. Calculate log₂ ratio: log₂Cy3 – log₂Cy5.

2.8 TaqMan (TAQ)

TaqMan® Gene Expression Assays are run in **four** replicates for each RNA sample. Four selected endogenous controls to be used for normalization are run on each 384-well plate: 18S (Hs99999901_s1), UBC (Hs00824723_m1), HPRT (Hs99999909_m1), and POLR2A (Hs00172187_m1).

Data will be analyzed using ΔC_T (absolute C_T value – endogenous control C_T value) and $\Delta\Delta C_T$ (ΔC_T for each gene – ΔC_T for a common reference gene) methods. The endogenous control with the least variation across all the samples and plates within one sample will be used for normalization to calculate $\Delta\Delta C_T$ for each individual gene. In the TAQ data submission, POLR2A was used as the internal control gene. TaqMan assay are provided in an Excel spreadsheet with each row representing an assay and each column representing a replicate. Please refer to the release note provided with TAQ data distribution for more information about the TAQ data format.

2.9 QuantiGene (QGN)

Three replicates are run for each of the 245 selected genes for each sample using the QuantiGene platform. The Relative Luminescent Unit (RLU) data are reported in an Excel file with each row representing an assay and each column representing a replicate.

For those genes (*e.g.* selected genes from MAQC pilot II study) whose expression in Ambion Brain and UHRR is substantially different (*e.g.* over 100:1 ratio), assay signal ratios of 100:75:25:0 for the four total RNA samples are expected. The closeness to the expected 100:75:25:0 ratios will be a measure of the accuracy of a platform. Please refer to the release note provided with QGN data distribution for more information about the QGN data format.

2.10 StaRT-PCR (GEX)



1. **Three** replicate measurements for each gene are obtained for each of the four mRNA samples. The loading control genes will be ACTB and GAPD.
2. *StaRT-PCR*TM transcript abundance measurements will be obtained using the Caliper AMS90TM data by analyzing the area under the curve. Data will be registered and automatically analyzed using the HT S-GEM SuiteTM software program. The data will be exported into a text file.
3. *StaRT-PCR*TM data are automatically processed and analyzed as described in SOP0009. Quality control specifications include appropriate negative and positive control wells on each 96 well plate. Peaks must meet a minimum peak height and peak area requirement. Both Native and Internal Standard peak base pair sizes must lie within a maximum designated range of ± 35 base pairs.
4. The final data report will contain TA measurements for the three individual replicates, the average, the standard deviation, and the coefficient of variation. The final report will be sent directly to the MAQC via electronic file.
5. Data on 207 genes were provided. Please refer to the release note provided with GEX data distribution for more information about the GEX data format.

2.11 Summary of Platform-specific Data Treatment

Each platform manufacturer will provide a description of its preference on how its data should be handled for consistent treatment regarding

1. Background subtraction
2. Normalization
3. Transformation
4. Data filtering
5. QC cutoffs for flags
6. QC cutoffs for failed samples
7. Platform-specific issues

Table 4. Summary of Platform-specific Data Treatment

No.	Platform	Background Subtraction	Normalization Method	Flags
1	AFX	Location-dependent	PLIER+16	P
2	AGL	Local?	Linear Lowess; dye-swap averaging	P
3	ABI	Global	Quantile	SNR \geq 3
4	EPP	Mean local; Mean of negative controls	Internal standards and house-keeping genes	Quantitative
5	GEH	Local median	Median of 1	G
6	ILM	Global	Quantile	Detection \geq 0.99
7	NCI	Local median	Lowess	0
8	TAQ		ΔC_T	
9	QGN		Relative Luminescent Unit (RLU)	
10	GEX		1,000,000 ACTB molecules	

2.12 Unique Probe ID Column(s) for Each Platform

Table 5. Summary of Platform-specific Unique/Preferred Probe(set) ID Column(s)

No.	Platform	Probe ID*	Example	ArrayTrack Data Field
1	AFX	Probe Set ID	1007_s_at	GEN_ID_MFR
2	AGL	ProbeName	A_23_P80353	GEN_ID_MFR
3	ABI	ProbeID	100002	GEN_ID_MFR
4	EPP	Gene ID	ABI2	GEN_ID_MFR
5	GEH	Probe_name**	GE766244	GEN_ID_MFR
6	ILM	TargetID	GI_10047089-S	GEN_ID_MFR
7	NCI	oligo_id	H200000226	GEN_ID_MFR
8	TAQ	Assay ID	Hs00605099_m1	GEN_ID_MFR
9	QGN	Catalog	PA-10181	GEN_ID_MFR
10	GEX	AssayID	BZW1	GEN_ID_MFR

*The Probe ID column shown in the original data files or the annotation files provided by the platform provider.

** GE Healthcare's preferred probe ID: "Probe_name" (e.g., GE766244). It is mapped to ArrayTrack's database field "GEN_ID_MFR" in the distributed "norm" data. "Feature_id" appears to be the only unique ID column that allows for cross-lab data mapping for the GEH platform; it is mapped to ArrayTrack's data field "POS_DESCR_MFR". For the "Probe_name" of "GE766244", the corresponding "Feature_id" is "1002". There are duplicates for some of the "Probe_name" entries in the data files. *Nonunique Probe_names on the GEH arrays correspond to positive and negative control spots and fiducials. All of the Discovery "probe types" on the GEH arrays have unique Probe_names.*

2.13 Manufacturer's Contact Person

If you have any general questions regarding the MAQC data distribution and data analysis, please contact Leming.Shi@fda.hhs.gov. For platform-specific questions, please contact the platform provider (and cc me):

ABI: Yongming Sun (650-638-6879, sunya@appliedbiosystems.com)
AFX: Xu Guo (408-731-5658, xu_guo@affymetrix.com)
AGL: Jim Collins (650-485-5048, jim_collins@agilent.com)
EPP: Françoise de Longueville (+32-81725615, delongueville.f@eppendorf.be)
GEH: Rich Shippy (480-722-2339, Richard.Shippy@ge.com)
ILM: Shawn Baker (858-202-4597, scbaker@illumina.com)
NCI: Ernie Kawasaki (301-435-2891, kawasake@mail.nih.gov)
TAQ: Kathy Lee (650-554-3496, Kathy.Y.Lee@appliedbiosystems.com)
QGN: Yuling Luo (510-818-2623, ylo@genospectra.com)
GEX: Jim Willey (419-383-3541, jwilley@meduohio.edu)

3. Data Management with ArrayTrack

ArrayTrack, a software package developed by the FDA/NCTR, is being used by the FDA to manage, analyze, and interpret Voluntary Genomic Data Submissions (VGDS). The MAQC

datasets are being managed by using ArrayTrack. It is anticipated that VGDS data will be submitted to the FDA in different platforms. Therefore, the FDA/NCTR is making every effort to make sure that ArrayTrack is able to handle data from multiple platforms such as those used in the MAQC main study. Although most manufacturers' preferred data preprocessing and normalization methods have been implemented within ArrayTrack, we decided to let the manufacturers provide normalized data for direct input to ArrayTrack in order to avoid potential inconsistencies in software implementations.

3.1 Annotation Information Provided by Manufacturers

Updated annotation files are available from platform providers or their web sites. Those who are interested in obtaining copy of the manufacture-provided annotation files may contact Leming.Shi@fda.hhs.gov.

3.2 Data Columns Imported to ArrayTrack

AFX:

- CEL files;
- Call justplier (Bioconductor) from ArrayTrack to generate PLIER output;
- Add an offset value of 16.

AGL:

The normalized data columns from Agilent's Feature Extraction software were directly input to ArrayTrack:

- gProcessedSignal (FEATURES Column 21) – normalized green intensity
- rProcessedSignal (FEATURES Column 22) – normalized red intensity
- combined flag (the last column of the distributed original text files)

ABI:

Quantile normalized data were provided by ABI and directly input to ArrayTrack:

- Signal
- S/N ratio

EPP:

Two-stage normalized data were provided by EPP and directly input to ArrayTrack:

- Relative intensity normalized to a reference sample
- STATUS (Quantitative, Qualitative, and Not in linear range)

GEH:

Median-scaling (to 1 per array) normalized data were provided by GEH and directly input to ArrayTrack:

- Normalized_intensity
- Quality_flag (CODELINKFLAG)

ILM:

Quantile normalized data were provided by ILM and directly input to ArrayTrack:

- AVG_Signal
- Detection

NCI:

- F635 Median (Cy5 spot intensity)
- B635 Median (Cy5 background)
- F532 Median (Cy3 spot intensity)
- B532 Median (Cy3 background)
- Flags (Flags, “0” indicates good spots; a negative flag value indicates a unreliable spot)
- The Lowess algorithm implemented within ArrayTrack was applied.

TAQ:

The sign of the ΔC_T (absolute C_T value – endogenous control C_T value) values provided by TAQ was reversed so that a larger value would represent a higher expression level. That is, the data input to ArrayTrack (and distributed in the “norm” file) is the value defined as ΔC_T (endogenous control C_T value – absolute C_T value).

QGN:

Background-corrected RLU signal
Flag

GEX:

Number of molecules (normalized to 1,000,000 ACTB molecules)

3.3 Data Distribution for Independent Analysis

For each platform, the data distribution includes two components: normalized data and original data.

Normalized data:

Normalized data were exported from ArrayTrack in a consistent format including the following columns:

- GEN_ID_MFR column that corresponds to the manufacturer’s unique probe(set) ID as shown in Table 4.
- SPOTID column that is a unique identifier within ArrayTrack. Analysis sites can simply ignore it.
- log₂ transformed normalized signal intensity values.
- Flag information provided by the manufacturer.

Notes:

- In order to avoid confusion in cross-platform comparison, I have deliberately ignored other annotation information (*e.g.*, gene ID, gene description, RefSeq, Locus ID, *etc.*) provided by the manufacturers. The sequence-based cross-platform mapping index file being created by Damir Herman and Rick Jensen will be made available for cross-platform data comparison.
- The 60 arrays from the three “official” test sites were exported into one single tab-delimited text file.
- A brief “release note” was provided for each platform.
- A simple hierarchical cluster figure (heatmap) was provided for a quick overview of the data quality.
- All the above files were WinZipped into one single file (*e.g.*, norm_AFX_123_PIER16.zip) that indicates the form of the data (norm for normalized), platform (AFX), test sites (1, 2, and 3), and normalization method (PIER+16).

Original data:

Original data were provided in a single WinZipped file (*e.g.*, ori_AFX_123_CELs.zip) that indicates the form of the data (ori for original), platform (AFX), test sites (1, 2, and 3), and the original data format (CEL files).

All zipped files are password-protected; call Leming Shi at 870-543-7387 for the password. Please contact Leming.Shi@fda.hhs.gov if you need to access image files. More detailed instructions on data distribution have been provided to data analysis sites and test sites that have responded to the “Conditions for Accessing and Analyzing the MAQC Datasets” sent out on October 24, 2005.

4. Microarray Quality Control Metrics and Thresholds

One important goal of the MAQC study is to assess the best performance that can be achieved by microarray technology under consistent experimental conditions so that end users will be able to judge whether the quality of their microarrays is comparable to the achievable performance of the platform. In doing so, procedural failures of a laboratory or operator may be identified and avoided.

To assess the performance of a technology it is imperative to define a set of parameters (or metrics) that evaluate the quality from different perspectives. In this section, we intend to enlist a set of metrics that have been widely used and generally accepted by the microarray community in one way or another to assess microarray quality. It should be noted that some of these QC metrics are platform-dependent and may not apply to all platforms involved in the MAQC study. In addition, it may not be feasible for each platform provider or data analysis site to analyze the MAQC datasets using all the QC metrics mentioned in this section. It is also expected that the MAQC team may propose additional QC metrics.

It should be noted that it is not the objective of the MAQC project to rank different microarray platforms.

4.1 QC Metrics Based on a Single Array

Single-array based QC metrics can be derived from a single array and does not rely on other arrays in the same study. They can be categorized into two types: before and after hybridization.

Single-array QC Metrics (Before Hybridization)

Before hybridization, single-array QC metrics reflect the quality of the individual steps in the microarray analysis process before hybridization:

1. RNA A260/A280 ratio (purity)
2. RNA Concentration (ug/uL)
3. RNA 28S/18S ratio (integrity)
4. RIN (RNA Integrity Number; integrity)
5. cDNA or cRNA Yield (ug)
6. cDNA or cRNA median fragment size (NTs)
7. cDNA or cRNA Labeling efficiency (dye molecule/NTs)
8. Platform-specific QC (??)

where metrics 1-4 refer to the quality of the total RNA sample (and do not depend on the performance of a microarray platform) and metrics 5-7 refer to the quality of target preparation.

Single-array QC Metrics (After Hybridization)

After hybridization, single-array QC metrics reflect the quality of an individual hybridization:

9. Image defects (qualitative, visual inspection)
10. Median signal-to-noise ratio (SNR)
11. Median feature intensity
12. SD of feature pixel intensity
13. SD of intensity for replicate features
14. SD of log₂ ratio of replicate features for 2-color arrays
15. Spot morphology
16. Number of unreliable features (saturation, not-found, *etc.*)
17. Spike-in performance
18. 3'/5' GAPDH ratio
19. Global 3'/5' ratio
20. % of present calls
21. Platform-specific QC (??)

4.2 QC Metrics Based on Replicate Arrays

The use of replicate arrays allows for the assessment of a platform's measurement precision, which is an important feature of a measurement technology. The following metrics measure the consistency in replicate arrays:

22. Consistency of % present calls
23. Consistency of genes being called present
24. Intensity correlation
25. log intensity correlation
26. False positive rate
27. Ratio correlation (*For one-channel platform, should we consider all the pairing possibilities? If so, for the comparison of RNA samples A and B, each with five replicates, there would be 25 possible pairs.*)
28. log ratio correlation
29. Percentage of overlapping genes (POG) from two significant gene lists
30. Platform-specific QC (??)

where metrics 22-25 refer to the consistency of intensity measurement for replicate arrays within the same RNA group and metrics 27-30 refer to the consistency of ratio (fold change) measurement for replicate array pairs in which the expression differences between two different RNA samples are measured.

4.3 QC Metrics Based on a Study

The quality of a study that compares the differential gene expression between two groups of RNA samples depends not only on the within-group data consistency but also on the between-group difference. When the between-group difference (*i.e.*, the true biological difference) is large, the detected fold change becomes more reliable. Higher within-group consistency (precision) also leads to more reliable detection. The following metrics define the overall quality of a microarray study in which two groups of samples are compared:

31. A “measure” to define the intrinsic difference between the two groups of samples:
$$\frac{\text{Mean within group correlation}}{\text{Mean between group correlation}}; \text{Mean within group correlation} - \text{Mean between-group correlation}; \textit{etc.}$$
32. *p*-value plot (ArrayTrack)
33. Gene Ontology
34. Pathways
35. Class Prediction
36. Others

Needs more thoughts on this section.

We cannot do anything if there is no difference between the two groups of samples in a study, or the difference is below the detection limit of microarrays.

--Thanks to Walter Liggett for the following suggestions:

In the FDA, the most important comparison of two groups of samples is a clinical trial. A paradigm for clinical trials helps put QC metrics in context. Richard Simon (NCI) has such a paradigm, which is summarized in the following slide from one of his talks:

The Paradigm

1. Develop a completely specified pharmacogenomic (PG) classifier of the patients likely to benefit from a new medical product (E)
2. Establish reproducibility of measurement of the classifier
3. Use the completely specified classifier to design and analyze a new clinical trial to evaluate effectiveness of E in the overall population or pre-defined subsets determined by the classifier

This paradigm can help with thinking about QC metrics even though Richard Simon's paradigm is still being discussed.

From the FDA perspective, the focus of QC metrics must be Step (2): Establish reproducibility. However, QC metrics are needed before the classifier is specified. They are needed to assure that the best measurements go into development of the classifier, and they are needed to provide some assurance that when step 2 is reached, it can be accomplished quickly. Thus, QC metrics general enough to cover most possible classifiers are needed. There are four types of such QC metrics:

1. QC metrics based on a distance that portrays the difference between the results from two microarrays hybridized with the same reference material. For a group of microarrays, this distance leads to a distance matrix, and this matrix can be portrayed with multidimensional scaling. The familiar principal components plot is actually an example of this.
2. QC metrics based on a collection of distances that lead to a collection of distance matrices. For example, one might compute a distance matrix for each of the MAQC reference materials and see if between-array differences have the same pattern for each reference material. Distance matrices can be compared through the use of Procrustes analysis.
3. QC metrics based on a statistic (such as a t statistic) computed from the expression values obtained for each gene in a group of genes. In the case of the t statistic, the result can be a volcano plot or one of several alternatives to the volcano plot.
4. QC metrics based on factor analysis. Factor analysis differs from principal components analysis in that it incorporates specific factors. Specific factors allow the noise level to vary from array to array.

These different types of QC metrics are all needed because they provided different views of the reproducibility of microarray measurements.

When a classifier has been completely specified, focus shifts to a QC metric for the particular classifier. Such a QC metric will typically have no more than a few dimensions, and the choice of the QC metric will be much more familiar. However, in a reproducibility study for a classifier, one will still want to look at performance more generally. One reason is that one will be looking for opportunities to improve the measurements underlying the classifier. Another reason is that one will want to prepare for eventual upgrading of these measurements on the basis of new technology.

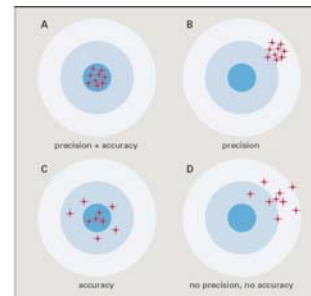
There are two uses of the MAQC data for which providing data analysis guidance seems especially important. One is the case in which a laboratory wants to check its performance by comparing its results on the MAQC reference materials with the results obtained in the MAQC study. Another is use of the MAQC data to compute the scores underlying a particular classifier. Such computation should provide some indication of the reproducibility of the measurement of the classifier even though the MAQC reference materials are neither cases nor controls.

-- End of quotation from Walter.

4.4 QC Metrics on Accuracy

It is important to differentiate precision from accuracy in assessing the performance of a measurement technology like microarrays. Precision refers to the closeness of repeated measurements to each other, and accuracy indicates the closeness of the measured result to the truth. It is a common practice in analytical chemistry to use reference or standard

materials alone or in a spike-in (or standard addition) manner to assess the accuracy of an analytical procedure for determining the substances of interest. Microarrays are a technology that tries to simultaneously determine tens of thousands of different molecules (transcripts). It would be ideal to have reference or standard materials for each of the transcripts to evaluate the performance of a microarray platform. Unfortunately, making standard materials for each transcript is a daunting task and beyond the reach of today's technology.



(Figure source: http://www.affymetrix.com/support/technical/technotes/platforms_technote.pdf)

Alternatively, external spike-in control transcripts are added to the test samples for some platforms to monitor the performance of various steps such as reverse transcription, labeling, and hybridization. In addition, titration of two RNA samples at different mixing proportions has been used for assessing accuracy. Since microarray data have been routinely “validated” by a more widely accepted technology (*e.g.*, QRT-PCR), it is natural to use the consistency between microarray measurements and those of QRT-PCR as a surrogate of accuracy measure. If we assume that each microarray platform is able to detect a portion of true answers (*e.g.*, differentially expressed genes between two RNA samples) and that each platform fails to detect the truth in its own unique way, the consistency across multiple microarray platform may also be used as a surrogate of accuracy measure, and a platform showing a better overall consistency with other platforms may be considered having a better accuracy.

37. Performance of spike-in controls (in terms of intensity and ratio)
38. Expected versus measured ratios for a set of sample-specific genes in the titration mixtures
39. Comparability with TaqMan assays
40. Consistency across multiple microarray platforms

4.5 QC Metrics for Cross-site Comparison

The QC metrics listed in Sections 4.1 to 4.4 are used for assessing the performance of the same platform in the same laboratory by the same operator, as has been designed in the MAQC main study, and many of them can be easily adapted to assess the consistency of data generated on the same platform by different test sites. Cross-site consistency is a good indication of the robustness of a platform.

4.6 QC Metrics for Cross-platform Comparison

The QC metrics listed in Sections 3.1 to 3.4 are used for assessing the performance of the same platform in the same laboratory by the same operator, as has been designed in the MAQC main study, and some of them can be adapted to assess the consistency of data generated on different platforms by different test sites.

4.7 QC Thresholds

Defining a set of QC metrics is important for microarray quality control and quality assurance. However, a QC metric without an appropriate threshold (cutoff) is useless. Therefore, a critical component of the MAQC project is to establish the threshold for each QC metric from the large datasets generated by the MAQC so that end users can easily know whether the performance of their laboratories or operators is good enough. Within the MAQC project, we are trying to establish the achievable performance of each microarray platform. However, whether the achievable performance of a platform is good enough for a particular application is out of the scope of the current study, and more research is needed to address it.

Here is an example for establishing the threshold for log intensity correlation for a platform in the same test site (illustrated using one-color platform). Each of the four RNA samples is tested in five replicates. Therefore, we can calculate 10 pair-wise correlation coefficients for the five replicates of each RNA sample. Statistics (*e.g.*, mean and SD) can be calculated using the 40 pair-wise (within replicates) correlation coefficients from the four samples. Alternatively, the calculation of mean and SD can be extended to the 120 pair-wise correlation coefficients from the three test sites. The Mean \pm SD may be used as the threshold.

Similarly, the threshold for log ratio correlation can be established. However, it is likely that the log ratio correlation is dependent on the intrinsic biological differences between the pair of RNA samples being compared. For the four RNA samples (Table 1), there are six possible ways of pairing (A-B, A-C, A-D, B-C, B-D, and C-D). Therefore, the threshold for log ratio correlation may need to be established separately for each pairing. Note that end users outside of the MAQC project will only have access to samples A and B; thus the threshold for log ratio correlation for A-B pair is more important.

5. Criteria for Defining Subsets of Genes for Cross-platform Comparison

The RefSeq database has been used by many manufacturers as the major source of target sequence information to establish the probe-target relationship. Consequently, a decision was made during the 2nd MAQC face-to-face meeting in Rockville, MD (May 2-3, 2005) to use RefSeq matching as the basis for cross-platform comparison. The main challenge is to identify a subset of common RefSeq transcripts that different platforms are trying to detect. Different lists of common RefSeq IDs will be generated based on different criteria in determining whether two platforms are indeed measuring the same transcript.

5.1 Matching of RefSeq IDs Based on Manufacturers' Annotation

Based on the annotation information provided by each manufacturer, a subset of RefSeq IDs (NM_#) is identified and used for cross-platform comparison. For simplicity, it is assumed that for every platform each probe (set) is mainly used for measuring one transcript (RefSeq ID).

5.2 Mapping of Probe Sequences to RefSeq Transcripts

Damir Herman has BLASTed each platform provider's probe sequences against the RefSeq database. It is up to the MAQC group to decide the cutoffs for determining whether two platforms are measuring the same transcript. It is possible that multiple lists of common RefSeq IDs will be generated based on different cutoffs.

Criteria for selecting RefSeq Transcripts:

1. The probe sequence perfectly matches a RefSeq transcript. Note that, after consulting with Affymetrix, we learned that at least nine of the ~11 PM probes must match to the same RefSeq transcript for a probe set to be considered as a perfect match to the transcript.
2. A shorter list can be generated by limiting to those RefSeq IDs for which each platform's probe targets to the 3'-end (*e.g.*, within 1,500 bases from the first base of the probe sequence to the 3'-end).
3. We can further exclude RefSeq IDs for which some platform's probe may show strong cross-hybridization to another transcript.
4. Other criteria may be proposed the MAQC participants.

5.3 Handling of Probe-Transcript (Target) Relationships

There are several scenarios regarding the probe-transcript relationships (1-4). The MAQC group needs to decide how to deal with the different probe-transcript situations. (Note: The notations are based on a technical presentation of GE Healthcare.)

1. One Probe – One Transcript (Single)

The probe is specific to one known transcript. This is the most common scenario. *All probes in this category should be included.*

2. Multiple Probes – One Transcript (Duplicate)

This may be a result of the dynamic change of public sequence databases or intentional design by the manufacturer. *Should we use some sort of averaging to represent the measurement for the transcript? We might want to look at the probes separately and then average if the results of the various probes for the same transcript are similar.*



3. One Probe – Multiple Transcripts (*Multiple*)

There are many possible reasons for this to happen. The probe may target more than one alternative transcript for a locus. The probe may target more than one member of a paralogous gene family. UniGene data is dynamic and the cluster may have been split into two or more distinct clusters, and the probe may now target more than one cluster. *Should we ignore all the transcripts or use the single measurement value of the probe to represent the expression for multiple transcripts? This information will be valuable to the end-users because these probes are not detecting unique transcripts but the sum of multiple (some related and some unrelated) transcripts. The end-user may need this information to help interpret the microarray results. I hope the manufacturers will agree to make this information public.*

4. Multiple Probes – Multiple Transcripts (*Multiple*)

This is a combination of the above (A-C). *What should we do, ignore them altogether?*

There have been many discussions between the sequence-based mapping group and manufacturers on probe-target mapping. Some degree of consensus has been reached. As suggested by Shawn Baker, master index files in the format of Table 6 to map each platform’s probes to the RefSeq database have been generated by Damir Herman and Rick Jensen to facilitate cross-platform data comparison. Different versions of this table are expected to be generated when different sets of “rules” are applied.

Table 6. The NM Table

RefSeq Transcript ID	Platform I	Platform II	...	Platform N
NM_...	Matching Probe ID	Matching Probe ID	...	Matching Probe ID
....				
NM_...				

Notes from Shawn Baker (thanks, Shawn):

- Begin with probe sequences from each platform and transcript sequences from RefSeq.
- For each platform, filter out any probes that are not valid. A probe is defined as valid if it perfectly matches a transcript sequence and does not perfectly match any other transcript sequence with a different gene symbol¹.
- For each transcript sequence and each platform, determine if the transcript sequence perfectly contains a valid probe. If a transcript sequence contains multiple valid probes, select the one closest to the 3’ end of the transcript².
- Generate an output table like the one below. Some cells in the table will be blank in the case that a transcript is not covered by a valid probe of a particular platform.

¹ For platforms that contain multiple probes per probe set, proceed as follows: Determine which probes are valid by applying the definition above, and define a probe set as valid if at least 80% of its probes are valid.

² Choosing at most one probe per transcript from each platform will facilitate downstream analysis. It is not clear how to do a comparison across thousands of transcripts if platforms have differing numbers of probes per transcript.

5. Contact Person for the Probe–Target Mapping Efforts

Please contact Damir Herman (herman@ncbi.nlm.nih.gov, 301-594-2274) and Rick Jensen (roderick.jensen@umb.edu, 617-287-6032) for questions regarding to the probe-target mapping efforts and for accessing different versions of the “NM” table.

5.4 Handling of Flags and Low-intensity Transcripts

The complexity of microarray data analysis begins as soon as the image is collected. Each manufacturer has developed its own way of handling low intensity spots and assigning a flag status to each spot to indicate the reliability of the measured value. Two extreme cases are:

1. Do not filter any data. Consider each feature as a potentially useable data point.
2. Only consider those features that are flagged as “0” (“Present” or “good”) for all the samples being compared.
3. Strategies between the two extremes. *How?*

6. Data Analysis Considerations

6.1 Outlier Arrays

It is expected that a (small) portion of the replicate arrays will fail and become obvious outliers. It is reasonable to exclude such outlier arrays from further data analysis. *However, how to determine outliers without biases? We should also publish how many outliers there are at each site for each platform. That will give some real-world information on the use of the microarrays and what kind of consistency can be obtained - very useful to end-users. Failure rate = ???*

6.2 Detection of Expressed Genes - Present Calls

Each platform provider has set its own criteria for making present or absent calls, and such information is usually available from the output file for each array. If three or more replicates out of the five detect a transcript as present, then this transcript will be considered as present by the platform in the test site.

6.3 Detection of Differentially Expressed Genes - Gene Selection Methods

There are many ways to identify genes that are significantly differentially expressed between two samples under comparison. Several commonly used methods include:

1. Simple fold change (ratio) cutoff (*e.g.*, 2-fold)
2. Single *p*-value cutoff from a statistical test (*e.g.*, $p \leq 0.01$)

3. A combination of fold change and p -value cutoffs (e.g., 2-fold and $p \leq 0.01$)
4. SAM (Significance Analysis of Microarrays)
5. False Discovery Rate (FDR, 5%)
6. Others

6.4 Normalization Methods

I am hoping that (in the future) we do not have to do any normalization at all! I would also be satisfied if a common set of ERCC spike-in controls will be demonstrated to be good for normalization.

The diversity and impact of different ways for pre-processing and normalizing microarray data reflect another major challenge to the microarray community and regulatory agencies. The impact of the following normalization/pre-processing procedures will be investigated according to some of the QC metrics mentioned in this document:

1. Original data (no normalization)
2. Total (or mean) intensity scaling
3. Median intensity scaling
4. Quantile normalization
5. Affymetrix-specific:
 - MAS
 - PLIER
 - RMA
 - GCRMA
 - dCHIP
 - Many other variants

Additionally, the impact of different ways for background handling will also be investigated.

Note: For multi-array based normalization methods (e.g., Quantile and PLIER), the 20 arrays from a test site will be treated as an independent group to generate the normalized expression values.

6.5 Exploratory Analysis Methods

Numerous methods can be applied for analyzing microarray data. Some of the widely used approaches are listed:

1. Principal component analysis
2. Hierarchical cluster analysis
3. ANOVA
4. Others

It should be noted that such analyses can be performed with expression values (intensity) or differential expression values (ratio).

6.6 Correlation between Array Data Quality and the Quality of RNA, Target, etc.

Is there a good correlation?

7. Comparison of Microarray with TaqMan, QuantiGene, and StaRT-PCR

Comparison will be conducted on the log₂ ratio space since intensity values are generally not comparable across platforms and technologies. The comparison will be based on the common subsets of genes identified in previous sections between the platforms being compared.

- Microarray versus TaqMan
- Microarray versus QuantiGene
- Microarray versus StaRT-PCR
- TaqMan versus QuantiGene
- TaqMan versus StaRT-PCR
- QuantiGene versus StaRT-PCR

8. Analysis of Titration Datasets

Rich/Shawn will coordinate the analysis and publication of the Titration Datasets:

- Identification of Sample-specific Genes
- Expected Ratio versus Measured Ratio
- Differentially Expressed Genes
- Reliability of detecting small fold changes
- PCA
- HCA

Genospectra's description on using titration data to assess accuracy (thanks to Yuling Luo):

Because Sample C and Sample D are prepared from Sample A and Sample B, the amount of the mRNA present in Sample C and Sample D can be calculated based on Sample A and Sample B. If the amount of a target mRNA present in Sample A, B, C, D is denoted R_A , R_B , R_C , R_D , respectively, then

$$R_C = R_A * 0.75 + R_B * 0.25$$

$$R_D = R_A * 0.25 + R_B * 0.75$$

Accuracy Estimation:

The assay signal obtained for a target mRNA is denoted as S_A , S_B , S_C , S_D in Sample A, Sample B, Sample C, Sample D, respectively. If we assume that the assay signal for the target mRNA is within the linear dynamic range of the assay, then the predicted assay signal for Sample C and Sample D can be calculated using following formula:

$$S'_C = (S_A * 0.75 + S_B * 0.25)$$

$$S'_D = (S_A * 0.25 + S_B * 0.75)$$

The variation between the predicted signal S'_C and S'_D and the actual assay signal S_C and S_D could be used as an indicative parameter of assay accuracy.

An accuracy parameter δ_C and δ_D for a target gene can be defined as:

$$\delta_C = | (S_C - S'_C) | / S'_C$$

$$\delta_D = | (S_D - S'_D) | / S'_D$$

An overall accuracy parameter P can be defined as:

$$P_C = (\sum \delta_C) / N$$

$$P_D = (\sum \delta_D) / N$$

Where N is the total number of target genes measured.

It should be noted that this calculation is independent of the assay background signal for a target mRNA as long as the assay background signal is equal among all samples. Assuming B is the assay background signal, then

$$S'_C = (S_A + B) * 0.75 + (S_B + B) * 0.25 = (S_A * 0.75 + S_B * 0.25) + B$$

$$S'_D = (S_A + B) * 0.25 + (S_B + B) * 0.75 = (S_A * 0.25 + S_B * 0.75) + B$$

However, if assay signal for a target mRNA has a nonspecific cross-hybridization component, then the nonspecific cross-hybridization component is likely to be different among different samples, and this nonspecific cross-hybridization component is likely to affect the overall accuracy of the assay for the target mRNA. Assuming N_A and N_B represent the nonspecific hybridization contribution to the target mRNA in Sample A and Sample B, then it is clear that

$$S'_C = (S_A + B + N_A) * 0.75 + (S_B + B + N_B) * 0.25 = (S_A * 0.75 + S_B * 0.25) + (N_A * 0.75 + N_B * 0.25) + B$$

$$S'_D = (S_A + B + N_A) * 0.25 + (S_B + B + N_B) * 0.75 = (S_A * 0.25 + S_B * 0.75) + (N_A * 0.25 + N_B * 0.75) + B$$

In summary, assay accuracy can be measured by comparing the predicted assay signal for Sample C and Sample D to their actual assay signal. The assay accuracy will be affected when the assay signal is not within the linear dynamic range of the assay or when non-specific cross-hybridization occurs.

---- End of Genospectra's description

9. Performance of Spike-in Controls

Many platforms used in the MAQC main study include platform-specific spike-in controls. The performance of the spike-in controls will be investigated in order to provide preliminary data for implementing the ERCC testing plan. Performance will be examined in terms of intensity and ratio.

Spike-in controls are included for the following platforms:

AFX

AGL

EPP

...

10. Comparison of Scanner Performance

Each of the arrays from AGL, GEH, and NCI platforms are scanned on at least two brands of scanners. The data from different scanners are to be compared. Data collected from different brands of scanners with scanner calibration slides will also be compared.

11. Identification of Genes Showing the Largest Discordance

- Define a set of measures of “discordance”
- Select 20~30 genes for follow-up investigation (AB, Gene Express, Genospectra, and Stratagene)

12. Presentation of Analysis Results at the December 1-2, 2005 Meeting

Please see the tentative meeting Agenda.

13. Planning for Publications

There is high expectation from the MAQC group to have the results published in prestigious journal(s). Because of the huge efforts of the MAQC project and its potential impact to the microarray community, it would be ideal to identify a prestigious journal that is willing to publish the MAQC results in a special issue (monograph). A list of tentative topics for manuscript drafting is shown in Table 7.

14. References

Irizarry papers

Li-Wong dCHIP

Affymetrix presentations on PLIER

.....

1. Marshall E. Getting the noise out of gene arrays. *Science*, 306(5696), 630-631 (2004).
2. Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC. Evaluation of gene expression measurements from commercial microarray platforms. *Nucleic Acids Res*, 31(19), 5676-5684 (2003).
3. Larkin JE, Frank BC, Gavras H, Sultana R, Quackenbush J. Independence and reproducibility across microarray platforms. *Nat Methods*, 2(5), 337-344 (2005).
4. Irizarry RA, Warren D, Spencer F, Kim IF, Biswal S, Frank BC, Gabrielson E, Garcia JG, Geoghegan J, Germino G, Griffin C, Hilmer SC, Hoffman E, Jedlicka AE, Kawasaki E, Martinez-Murillo F, Morsberger L, Lee H, Petersen D, Quackenbush J, Scott A, Wilson M, Yang Y, Ye SQ, Yu W. Multiple-laboratory comparison of microarray platforms. *Nat Methods*, 2(5), 345-350 (2005).
5. Bammler T and Members of the Toxicogenomics Research Consortium. Standardizing global gene expression analysis between laboratories and across platforms. *Nat Methods*, 2(5), 351-356 (2005).
6. Shippy R, Sendera TJ, Lockner R, Palaniappan C, Kaysser-Kranich T, Watts G, Alsobrook J. Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics*, 5(1), 61 (2004).
7. Petersen D, Chandramouli GV, Geoghegan J, Hilburn J, Paarlberg J, Kim CH, Munroe D, Gangi L, Han J, Puri R, Staudt L, Weinstein J, Barrett JC, Green J, Kawasaki ES. Three microarray platforms: an analysis of their concordance in profiling gene expression. *BMC Genomics*, 6(1), 63 (2005).
8. Carter SL, Eklund AC, Mecham BH, Kohane IS, Szallasi Z. Redefinition of Affymetrix probe sets by sequence overlap with cDNA microarray probes reduces cross-platform inconsistencies in cancer-associated gene expression measurements. *BMC Bioinformatics*, 6(1), 107 (2005).
9. Mecham BH, Klus GT, Strovel J, Augustus M, Byrne D, Bozso P, Wetmore DZ, Mariani TJ, Kohane IS, Szallasi Z. Sequence-matched probes produce increased cross-platform consistency and more reproducible biological results in microarray-based gene expression measurements. *Nucleic Acids Res*, 32(9), e74 (2004).
10. Chen JJ, Delongchamp R, Tsai CA, Hsueh HM, Sistare F, Thompson KL, Desai VG, Fuscoe JC. Analysis of variance components in gene expression data. *Bioinformatics*, 20(9), 1436-1446 (2004).
11. Frueh FW, Huang SM, Lesko LJ. Regulatory acceptance of toxicogenomics data. *Environ Health Perspect*, 112(12), A663-A664 (2004).
12. Frueh FW, Goodsaid F, Rudman A, Huang S-M, Lesko LJ. The need for education in pharmacogenomics: a regulatory perspective. *Pharmacogenomics Journal* 5(4), 218-220 (2005).
13. Lesko LJ, Woodcock J. Translation of pharmacogenomics and pharmacogenetics: a regulatory perspective. *Nature Reviews Drug Discovery*, 3(9), 763-769 (2004).
14. Salerno RA, Lesko LJ. Pharmacogenomic data: FDA voluntary and required submission guidance. *Pharmacogenomics*, 5(5):503-505 (2004).
15. Shi L, Tong W, Fang H, Scherf U, Han J, Puri RK, Frueh FW, Goodsaid FM, Guo L, Su Z, Han T, Fuscoe JC, Xu ZA, Patterson TA, Hong H, Xie Q, Perkins RG, Chen JJ, and Casciano DA. Cross-platform comparability of microarray technology: intra-platform consistency and appropriate data analysis procedures are essential. *BMC Bioinformatics*, 6(Suppl 2), S12 (2005).



16. Shi L, Tong W, SU Z, Han T, Han J, Puri RK, Fang H, Frueh FW, Goodsaid FM, Guo L, Branham WS, Chen JJ, Xu ZA, Harris SC, Hong H, Xie Q, Perkins RG, Fuscoe JC. Microarray scanner calibration curves: characteristics and implications. *BMC Bioinformatics*, 6(Suppl 2), S11 (2005).
17. Shi L, Tong W, Fang H, Goodsaid F, Frueh FW, Han T, Fuscoe JC, and Casciano DA. QA/QC: Challenges and pitfalls facing the microarray community and regulatory agencies. *Expert Review of Molecular Diagnostics*, 4(6), 761-777 (2004).
18. The MAQC Project: Establishing QC Metrics and Thresholds for Microarray Quality Control, Abstract accepted for oral presentation at the MGED 8 Meeting, September 11-13, 2005, Bergen, Norway.
http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/docs/MAQC_MGED8Abs.pdf

15. Web Sites

1. The MAQC (MicroArray Quality Control) Project:
<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/index.htm>
2. ArrayTrack: <http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/>
3. Genomics at FDA: <http://www.fda.gov/cder/genomics/>
4. ERCC - <http://www.cstl.nist.gov/biotech/workshops/ERCC2003/>
5. ERRC - <http://www.cstl.nist.gov/biotech/ERCC/testplan.htm>
6. NIST Metrology for Gene Expression Program -
<http://www.cstl.nist.gov/biotech/Cell&TissueMeasurements/GeneExpression.htm>
7. Expression Analysis' Proficiency Testing using MAQC Samples -
http://www.expressionanalysis.com/proficiency_test.html
8. MAQC Sample A (Stratagene's Human Universal Reference RNA) -
<http://www.stratagene.com/manuals/740000.pdf>
9. MAQC Sample B (Ambion's Human Brain Reference RNA) -
<http://www.ambion.com/catalog/CatNum.php?6050>

The MAQC Project Meeting Agenda (Tentative)
December 1-2, 2005, Cyprus Room, Crowne Plaza Cabana, Palo Alto, CA

Thursday, December 1, 2005		
Chair: Yvonne Dragan (FDA/NCTR)		
8:00 am	Chair's remarks	Yvonne Dragan
8:05 am	Welcome	Ronald Davis (Stanford)
8:10 am	Overview of the MAQC project	Leming Shi (FDA/NCTR)
8:20 am	MAQC reference RNA samples	James Fuscoe (FDA/NCTR)
Seven Microarray Platforms* Chair: Federico Goodsaid (CDER)/Mike Wilson (Ambion)		
8:30 am	AFX: Inter-site reproducibility	Xu Guo (Affymetrix)
8:45 am	AGL:	Jim Collins (Agilent)
9:00 am	ABI:	Yongming Sun (Applied Biosystems)
9:15 am	EPP:	Francoise de Longueville (Eppendorf)
9:30 am	GEH:	Richard Shippy (GE Healthcare)
9:45 am	ILM:	Shawn Baker (Illumina)
10:00 am	NCI:	Ernest Kawasaki (NCI)
10:15 am	Break	
Three Alternative Technology Platforms* Chair: Federico Goodsaid/Mike Wilson		
10:30 am	TAQ (TaqMan):	Kathy Lee (Applied Biosystems)
10:45 am	QGN (QuantiGene):	Yuling Luo (Genospectra)
11:00 am	GEX (StaRT-PCR):	James Willey (Gene Express)
11:15 am Discussion (10 platform providers)	<ol style="list-style-type: none"> 1. Quality of data from your test sites 2. Lessons learned: challenges & pitfalls 3. Cost (\$ & time) and value to your org. 4. Suggestions for the MAQC project 5. Comments from external test sites 	Panel: Shawn Baker, Jim Collins, Francoise de Longueville, Xu Guo, Ernest Kawasaki, Kathy Lee, Yuling Luo, Richard Shippy, Yongming Sun, James Willey
12:00 pm	Lunch (on your own)	
Ten Data Analysis Sites (I)* Chair: Roderick Jensen (UMB)/Marc Salit (NIST)		
1:15 pm	Inter-laboratory variation	Walter Liggett (NIST)
1:30 pm	Correlation estimation from a random effects model	Sheng Zhong (UIUC)
1:45 pm	<i>Comparison of analysis methods</i>	Wenjun Bao/Tzu-Ming Chu (SAS)
2:00 pm	REDI	Wendell Jones (Expr. Analysis)
2:15 pm	<i>Data analysis pipelines</i>	Hanlee Ji (Stanford)
2:30 pm	VMAxS: Image-level data analysis	Cecilie Boysen (ViaLogy)
2:45 pm	QC metrics/thresholds & analysis methods	Leming Shi/Weida Tong (FDA/NCTR)
3:00 pm	Break	
Ten Data Analysis Sites (II)* Chair: Roderick Jensen/Marc Salit		
3:15 pm	Mapping probes to RefSeq and AceView	Damir Herman/Jean Thierry-Mieg (NCBI)
3:30 pm	Mapping probe sequences to RefSeq	Roderick Jensen (UMB)
3:45 pm	An extensive survey of analysis methods	Lisa Croner (Biogen Idec)
4:00 pm Discussion (10 analysis sites)	<ol style="list-style-type: none"> 1. Probe-target mapping & master index 2. QC metrics and thresholds 3. Normalization & gene selection 4. Dealing with flags 5. Future directions of data analysis 	Panel: Wenjun Bao, Lisa Croner, Damir Herman, Walter Liggett, Roderick Jensen, Hanlee Ji, Wendel Jones, Jean Thierry-Mieg, Weida Tong, Chunlin Xiao, Sheng Zhong
5:00 pm	Adjourn	

*Each presenter has 10 minutes for presentation and 5 minutes for questions/discussion.

Friday, December 2, 2005		
Chair: William Slikker, Jr. (FDA/NCTR)		
8:00 am	Chair's remarks	William Slikker, Jr.
8:10 am	Keynote address: TBA	Ronald Davis
Six Invited Presentations		Chair: William Slikker, Jr.
8:50 am	ERCC: The External RNA Controls Consortium	Janet Warrington (Affymetrix)
9:00 am	The Metrology for Gene Expression Program	Marc Salit (NIST)
9:10 am	Array standardization and the MfB Programme	Carole Foy (LGC, UK)
9:20 am	Study design and analysis of multiple platforms (mouse)	Winston Kuo (Harvard)
9:30 am	Proficiency tests using MAQC samples	Laura Reid (Expr. Analysis)
9:40 am	10 th Anniversary of <i>Science</i> 95' paper	Mark Schena (TeleChem)
Regulatory Perspectives:		Chair: William Slikker, Jr.
9:50 am	<ol style="list-style-type: none"> 1. Status of VGDS (Goodsaid/Tong presentation) 2. Expected uses of the MAQC datasets 3. Proficiency tests 4. External spike-in controls 5. Remaining challenges (prior to RNA) 6. Guidance on QC and data analysis 	Panel: Yvonne Dragan, Federico Goodsaid, Francis Kalush, Scott Pine, Hongzhu Ren, Karol Thompson, Weida Tong
10:30 am	Break	
Next Steps for the MAQC Project		Chair: Yvonne Dragan/Federico Goodsaid
10:45 am	MAQC for mouse and rat: Preliminary results (EPA and FDA) Survey of interests from platform providers RNA samples Experimental design and timeline	James Fuscoe Hongzhu Ren
	Trouble-shooting for discordant genes: List of discordant genes Follow-up experiments TAQ, QGN, GEX, and Stratagene	Federico Goodsaid Francis Kalush Scott Pine
Planning for MAQC Manuscripts**		Chair: Leming Shi/(Journal Editor, invited)
11:30 am	<ol style="list-style-type: none"> 1. Editorial 2. FDA's VGDS 3. MAQC main manuscript 4. Reference RNA samples 5. Sequence-based cross-platform mapping 6. Normalization & gene selection methods 7. Validation by alternative technologies 8. Titration datasets 9. Inter-laboratory variations 10. Cross-hybridization 11. One-color versus two-color platforms 12. Informatics tools for regulatory review 	Dan Casciano/Janet Woodcock Felix Frueh/Sponsors Leming Shi (MAQC Team) James Fuscoe/Mike Wilson Damir Herman/Roderick Jensen Leming Shi/Jim Chen Federico Goodsaid Richard Shippy Walter Liggett Zoltan Szallasi/Roderick Jensen Tucker Patterson/Jim Collins Weida Tong
12:55 pm	Concluding remarks	Leming Shi
1:00 pm	Adjourn	

**More detailed information will be provided in a separate document before the meeting.

Location:

The meeting will be held at the **Cyprus Room**, Crowne Plaza Cabana, 4290 El Camino Real, Palo Alto, CA 94306. Telephone: 650-857-0787; FAX: 650-496-1939; Website: <http://www.cppaloalto.crowneplaza.com/>. Please note the change of meeting venue.

Registration:

Registration is free but required. Contact Leming Shi, National Center for Toxicological Research, US FDA, 3900 NCTR Road, Jefferson, AR 72079. Tel: 870-543-7387; leming.shi@fda.hhs.gov.

Conference Call:

A teleconference will be set up to allow interested people to join the meeting by telephone throughout the entire MAQC meeting. Dial-in information will be provided when it is available. *It is also being explored to set up WebEx access for sharing meeting presentations.*

MAQC:

<http://edkb.fda.gov/MAQC/> or
<http://www.fda.gov/nctr/science/centers/toxicoinformatics/maqc/>.

ArrayTrack:

<http://www.fda.gov/nctr/science/centers/toxicoinformatics/ArrayTrack/>

Genomics @FDA:

<http://www.fda.gov/cder/genomics/>

Stanford Genome Technology Center (SGTC):

<http://med.stanford.edu/sgtc/>

ERRC (The External RNA Controls Consortium):

<http://www.cstl.nist.gov/biotech/ERCC/testplan.htm>

NIST's Metrology for Gene Expression Program:

<http://www.cstl.nist.gov/biotech/Cell&TissueMeasurements/GeneExpression.htm>

Co-sponsors: The MAQC (MicroArray Quality Control) project meeting is co-sponsored by the US Food and Drug Administration and Stanford Genome Technology Center of Stanford University.





Table 7. An outline of manuscripts proposed as a special journal issue on Microarray Quality Control (for discussion only)

No.	Tentative Title	Contents	Manuscript Team
1	Pharmacogenomics and the US FDA’s Critical Path Initiative to Medical Product Development: Technology-driven Innovations	<ul style="list-style-type: none"> • Editorial to this Special Issue on microarray quality control; • The importance of pharmacogenomics in medical product development and personalized medicine (as stated in the FDA’s Critical Path Initiative); • The reliability of technologies is essential to realize the promises of pharmacogenomics. 	<p>Dan Casciano (Director, FDA/NCTR) Janet Woodcock (Deputy Commissioner, FDA)</p>
2	Implementation of the FDA’s Voluntary Genomic Data Submissions (VGDS) Mechanism: Guidelines on microarray quality control and data analysis are essential	<ul style="list-style-type: none"> • Implementation of the VGDS mechanism at the FDA for sponsors to voluntarily submit genomic data (including data management, analysis, visualization, and interpretation; training of FDA reviewers; interactions between FDA and sponsors); • Lessons learned from several VGDS case studies; • Illustration of the challenges in the regulatory review of pharmacogenomic data: quality control of microarray experiments and guidance on data analysis. 	<p>Felix Frueh Federico Goodsaid Weida Tong Sponsors</p>
3	The MAQC Project: Establishing Calibrated RNA Samples, Reference Datasets, and QC Metrics/Threshold for MicroArray Quality Control (“MAQC Main Manuscript”)	<ul style="list-style-type: none"> • Overview of the MAQC project: Pilot-I; Pilot-II; Main Study (~1000 arrays); • Uniqueness and deliverables of the MAQC project; • Quality control metrics and thresholds for assessing the performance of each platform; • Microarray cross-platform comparability; • Concordance with alternative technologies (TaqMan, QuantiGene, and StaRT-PCR); • Correlation of step-by-step QC data with overall quality of microarray data; • Relative accuracy based on titration datasets; • Next steps for the MAQC project: mouse and rat; beyond total RNA. 	<p>Leming Shi Mike Wilson Uwe Scherf Laura Reid</p>
4	Establishing Well-calibrated Reference RNA Samples as an Efficient Quality Control Tool for Quantitative Gene Expression Profiling	<ul style="list-style-type: none"> • MAQC Pilot-I: Selection of two RNA samples as the reference materials using six objective criteria (based on data from 160 arrays on four microarray platforms): (A) Stratagene Universal Human Reference RNA and (B) Ambion Human Brain Reference RNA were selected; • Characteristics of the two RNAs: quality, purity, stability, availability, and reproducibility; • Characterization by microarray gene expression profiling; • Protocols on the applications of the reference RNA samples. 	<p>Jim Fuscoe Mike Wilson Gavin Fischer</p>
5	Probe Sequence Based Mapping across Microarray Platforms: “Are they measuring the same things?”	<ul style="list-style-type: none"> • Mapping probe sequences of each platform to the 9/26/2005 release of RefSeq database; • Statistics on the probe-target mapping; • Handling different mapping scenarios: (A) one probe – one target; (B) multiple probes – one target; (C) one probe – multiple targets; and (D) multiple probes – multiple targets; • Relationship between proximity of probes from different platforms and cross-platform consistency; • Web access to probe sequences and cross-platform mapping indices. 	<p>Damir Herman Rick Jensen Scott Pine Zoltan Szallasi</p>
6	Impact of Normalization and Gene Selection Methods on Microarray Studies	<ul style="list-style-type: none"> • Normalization methods: raw; total (mean) intensity; median intensity; quantile; Affymetrix specific – MAS5, dCHIP, RMA, PLIER, and their variants; • Gene selection methods: fold change, <i>p</i>-value, fold change + <i>p</i>-value, SAM, FDR, Others; 	<p>Leming Shi Xu Guo Rafael Irizarry</p>



		<ul style="list-style-type: none"> • Impact of the different combinations of normalization and gene selection methods on the stability of the “significant gene list” from the same platform; • Impact on microarray cross-platform comparability; • Identification of a reliable/stable “significant gene list”; • Impact of different ways of handling background and offset values. 	Wenjun Bao <i>Lisa Croner</i> Rick Jensen Jim Chen
7	Validation of Microarray Results with Alternative Technologies	<ul style="list-style-type: none"> • Unbiased selection of 1000 genes for TaqMan validation; • Unbiased selection of 200 genes for QuantiGene and StaRT-PCR validation; • Consistency across technology platforms; • Identification of the most discordant genes across platforms; • Design of new assays for explaining the discrepancies. 	Federico Goodsaid Kathy Lee Yuling Luo Jim Willey
8	The Use of Titration Datasets to Assess the Accuracy of Microarray Platforms	<ul style="list-style-type: none"> • MAQC Pilot-II: Selecting two of the 13 titration points for the main study (200 arrays plus TaqMan and QuantiGene assays for 10 genes); • MAQC Main Study: Two titration points (75:25 and 25:75, SUHRR:ABHRR); • Expected versus observed fold changes (including mathematical formulas); • The use of titration data to evaluate normalization methods. 	Rich Shippy Shawn Baker Paul Wolber Yuling Luo Rick Jensen
9	Characterization of Intra-platform, Inter-laboratory Variability	<ul style="list-style-type: none"> • Sources of technical variation that affect many genes; • Inter-laboratory comparison approaches that encompass many genes; • Surprising aspects of inter-laboratory differences; • Confirmation by manufacturer spike-ins (if this contributes to the characterization). 	Walter Liggett David Duewer Marc Salit
10	Modeling Cross-hybridization of Microarray Experiments	<ul style="list-style-type: none"> • Probe sequence based models to estimate the affinity between probe and target; • Probe sequence based models to estimate the level of tendency of cross hybridization; • Target sequence based methods to estimate the effect of secondary structure of microarray hybridization. 	Zoltan Szallasi Rick Jensen
11	One-color versus Two-color Designs in Microarray Experiments	<ul style="list-style-type: none"> • Philosophical considerations on microarray experimental designs: two-color vs one-color; • Common reference design vs sample-pairing design; • Dye biases and dye-swaps; • One-color hybridizations run on a two-color platform; • Ratio- and intensity-based data analysis for two-color design. 	Tucker Patterson Jim Collins Francoise de Longueville Hong Fang Tao Han Ernie Kawasaki
12	Informatics Tools for Utilizing the MAQC Datasets for Microarray Quality Control	<ul style="list-style-type: none"> • ArrayTrack as a tool for regulatory review of pharmacogenomic data submissions; • Assessing laboratory proficiency by comparing its data against the MAQC benchmark datasets; • Assessing the quality of individual arrays by examining the performance of spike-in controls. 	Weida Tong Hong Fang Federico Goodsaid